DOCUMENT RESUME

ED 044 120 LI 002 114

AUTHOR Standera, Oldrich
TITLE COMPENDEX/TEXT-PAC: CIS.
INSTITUTION Calgary Univ. (Alberta).
PUB DATE Aug 70
NOTE 94p.; Information Systems Report 6

EDRS PRICE EDRS Price MF-$0.50 HC-$4.80
DESCRIPTORS Costs, Information Centers, *Information Retrieval, *Information Science, *Information Systems, *Relevance (Information Retrieval)
IDENTIFIERS Canada, CIS, COMPENDEX, *Computerized Engineering Index, Current Information Selection

ABSTRACT

This report evaluates the engineering information services provided by the University of Calgary since implementation of the COMPENDEX (tape service of Engineering Index, Inc.) service using the IBM TEXT-PAC system. Evaluation was made by a survey of the users of the Current Information Selection (CIS) service, the interaction between the system and the users, system performance on the monthly computer runs, and user satisfaction. Costs of running time for some programs and of providing CIS, along with a pricing policy, are provided. A test of TEXT-PAC as a data base for full text searching capability was conducted, and the limiting factors of the TEXT-PAC system were explored. Because of its large data base and full text processing capability, the COMPENDEX system performance is reasonable. Improvements in relevance and recall can be made by improvements in formulating user profiles. Continuous system evaluation is required, along with user and search-editor training, improved feedback, and provision of word-frequency listings as an aid to correct profile formulation. (AB)

ED0 44120

THE UNIVERSITY OF CALGARY

DATA CENTRE

COMPENDEX/TEXT-PAC
CIS

by

OLDRICH STANDERA

August, 1970

INFORMATION SYSTEMS

Report No. 6

LI 002114

LI 001[redacted]

ED0 44120

U.S. DEPARTMENT OF HEALTH, EDUCATION
& WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

③ Calgary Univ. (Alberta), Data Center

THE UNIVERSITY OF CALGARY

DATA CENTRE

② all caps

COMPENDEX/TEXT-PAC:
CIS

by

① OLDRICH STANDERA

③ August, 1970

Report no.- INFORMATION SYSTEMS
Report No. 6

LI 002 114

CONTENTS

Page

1. Introduction 1
2. COMPENDEX 2
3. TEXT-PAC 5
4. Users of COMPENDEX in 1969 6
5. Interaction System-Users 11
6. The Monthly CIS Runs 15
7. Performance of the Service 18
    7.1 Relevance 19
    7.2 Analysis of Relevance 34
    7.3 Recall 38
    7.4 Analysis of Recall 42
    7.5 Precision-Recall 44
    7.6 "Miss" and "Trash" 46
    7.7 Comparison of AND, WITH, ADJ 49
    7.8 Match Criteria 1-3 50
    7.9 Searching Titles, Subject Headings and Abstracts . . . 52
8. Step Time of Some of the Programs 54
9. Calculation of the Cost of Current Information Selection . . 62
10. Pricing Policy 80
11. Input to TEXT-PAC other than COMPENDEX 84
12. Some Limiting Factors in the TEXT-PAC System 86
13. Conclusions 87

LIST OF FIGURES

Page

1. Scheme of TEXT-PAC (After Dr. S. Kaufman's TEXT-PAC Documentation) 4
2. Number of Search Expressions per Profile 9
3. Number of Profiles Processed 10
4. Profiles, Search Expressions, Words 11
5. Step Times of Programs 16
6. Monthly Runs 17
7. Relevance of Output (1969) 23
8. Feedback Received and Relevance (1969) 24
9. Dimensions in Indexing and Query Formulation 28
10. Exhaustivity, Specificity, Synonymity 29
11. Control of Relevance/Recall by AND, WITH, ADJACENT 31
12. Levels of Vertical Structure 32
13. Relevance Failures 36
14. Recall Values for Selected Profiles Output 41
15. Recall Failures 43
16. Relevance/Recall Graph 44
17. Relevance/Recall Table 44
18. Relevant/Irrelevant-Retrieved/Not Retrieved 46
19. "C" Evaluation (December, 1969) 47
20. AND, WITH, ADJ and Hits Received 50
21. Effect of MC on the Number of Hits 51
22. Title, Subject and Abstract Searching 53
23. Step Times (min), February, 1969; 1,527 abstracts 55
24. Step Time of the Profile Programs (10-70 Profiles) 56
25. Step Times of 360 Condensed Text Edit and Edit Convert--Varying Number of Documents 57
26. Relation between the Number of Profiles and CIS Search Step Time 58
27. Relation between the Number of Abstracts and CIS Search Step Time 59

Page

28. Step Times for 70 and 210 Profiles (CIS Search) 60
29. CIS Programs 66
30. Computer Cost (70 profiles) 67
31. Total Costs (70 profiles) 69
32. Cost per User, Profile, Search Expression, Word and Hit (70 profiles) 69
33. CIS Programs in Fig. 34 70
34. Computer Cost (210 profiles) 72
35. Cost per User, Profile, Search Expression, Word and Hit (210 profiles) 74
36. Time and Costs for 70 and 210 Profiles (TRC 011 CIS Search) 75
37. Cost of Individual Programs 76
38. Cost of SDI and Price per Profile ($) 77
39. Cost A through K (70 and 210 profiles) 78
40. Cost v. Price per Profile (70, 210, 280, 350 Profiles 79
41. SDI Price 82
42. TEXT-FAC Input Form 83

1. INTRODUCTION

The purpose of this Project Progress Report is to give an account of the work connected with the implementation of the COMPENDEX service using IBM's TEXT-PAC system, at The University of Calgary. In this report we are primarily concerned with the Current Information Selection (CIS). The experience gained in this work is applicable to the evaluation of other systems to be introduced on this campus.

CIS is more commonly known as Selective Dissemination of Information (SDI). Nowadays, SDI usually means a system where incoming documents are indexed or abstracted and processed into machine-readable form. Users' interest profiles are constructed and processed against the data base records.

From the above we can derive three major functions: abstracting, profiling and processing. These three functions may be done at one, two, or three organizations.

One of the essential features of any SDI system is the feedback from the user to the system. Its objective is to monitor the service to the user's satisfaction in terms of both relevance and recall.

As already mentioned we use the COMPENDEX data base of Engineering Index Inc., which is delivered in machine-readable form. Profiling is done both at The University of Calgary and AIRA, Edmonton. Machine-readable profiles are processed at The University of Calgary against the data base.

It was in April, 1969 that a recommendation was made to adopt the COMPENDEX service on this campus. The agreement between The University of Calgary and Engineering Index, Inc., is dated May 16, 1969. The actual work on the Compendex Project began in late June, 1969. The first data base tapes were processed in September, 1969.

Two persons were engaged in this project: one for preparing and adjusting profiles and input, evaluating the output and performance, for cost analysis, planning and directing the system; the other for computer operations, program control and submitting the jobs.

My thanks are due to Mr. Frank Dolan for his support and many fruitful discussions, and to Mr. Stan Nevlud for providing the interface with the IBM 360/OS.

2. COMPENDEX

COMPENDEX tapes are a service of the Engineering Index, Inc., United Engineering Center, 345 East 47th Street, New York, N. Y. 10017.

The data elements in COMPENDEX are arranged by means of the print controls as follows:

00b̸	Title 1st line
b̸b̸b̸	Title 2nd line to nth line
09b̸	Subject heading, subheading, EI Number
10b̸	Identification number
15b̸	CITE document accession number of items that are also part of CITE tapes
201	First author
202-299	Second author - 99th author
3b̸b̸	EI Number
4b̸Z	Citation
b̸b̸b̸	Citation - 2nd line to nth line
401	Author affiliate of 1st author
50b̸	Abstract - 1st line
b̸b̸b̸	Abstract - 2nd to nth line
60b̸	Subject heading, subheading
610 00-A to 649 00-A	Sales Codes (referring to EI card service) (CARD-A-LERT codes commencing summer 1970)
650-699	Access words
700	Source Index terms
750	Free language terms
95b̸	Table of contents (list authors and titles)
96b̸	Reserved

All of the print controls need not appear in the COMPENDEX files.

The input format is TEXT-PAC 360 condensed text. The maximum record length is 8004 bytes, variable length, unblocked. The magnetic tape is 9-track, 800 or 1600 BPI. The code used is Extended Binary Coded Decimal Interchange Code (EBCDIC). Tape length is 1200 feet.

Engineering Index Inc. is reviewing currently more than 3500 sources of engineering literature of all kinds and selected information is abstracted. Literature abstracted is stored in the Engineering Societies Library and is represented by professional, scientific and trade journals, publications of engineering organizations, associations, universities, laboratories and research institutions, government departments and agencies and industrial organizations, papers of conferences and symposia, selected books and patents.

The information in COMPENDEX tapes is pertinent to all of mechanical, chemical, electrical and civil engineering. The price of the tape is $500 monthly; if only one tape is ordered, the charge will be $750. The price of one reel is $25 charged extra.

The complete engineering information system consists of COMPENDEX tape service, the Engineering Index Monthly, the Engineering Index Annual. The purchase of the COMPENDEX tape service is contingent upon the subscription of both aforementioned indexes.

Engineering Index, Inc. also wants their customers to report

1. the number and kinds of clients,
2. pricing for this service,
3. fields of user interest,
4. to what extent the tape is being used,
5. for whom the service is being rendered,
6. what pricing and philosophy behind pricing,
7. value of service to the user,
8. any complaints or noise stemming from the service.

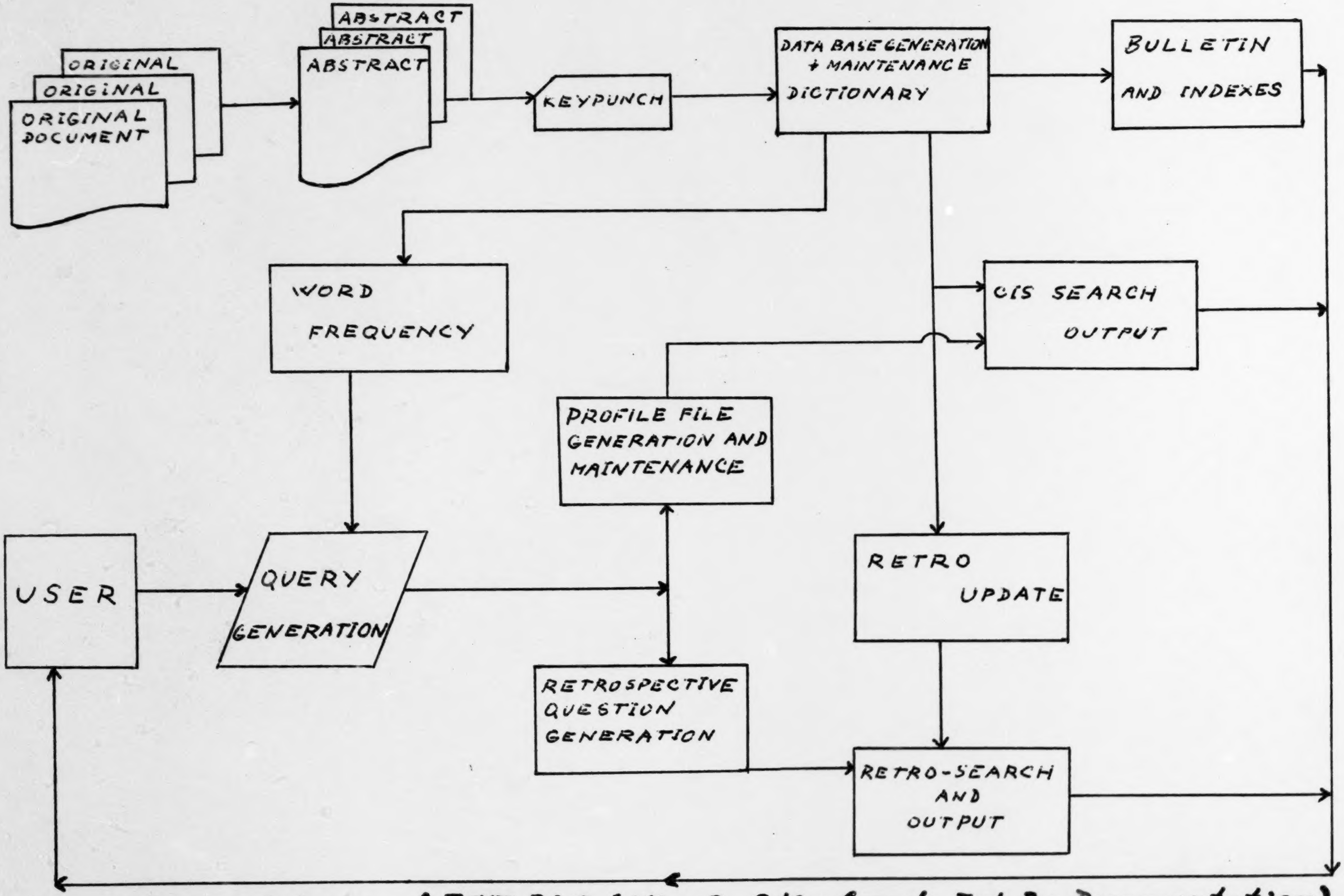

ORIGINAL
ORIGINAL
ORIGINAL DOCUMENT
ABSTRACT
ABSTRACT
ABSTRACT
KEYPUNCH
DATA BASE GENERATION + MAINTENANCE DICTIONARY
BULLETIN AND INDEXES
WORD FREQUENCY
CIS SEARCH OUTPUT
PROFILE FILE GENERATION AND MAINTENANCE
USER
QUERY GENERATION
RETRO UPDATE
RETROSPECTIVE QUESTION GENERATION
RETRO-SEARCH AND OUTPUT

Fig. 1. - Scheme of TEXT-PAC (After Dr. S. Kaufman's Text-Pac Documentation)

3. TEXT-PAC

The software for processing the COMPENDEX tapes is IBM's TEXT-PAC whose main author is Dr. Samuel Kaufman with A. V. Esposito, R. Fleischer, S. D. Friedman, S. Rogers, S. Skye, and U. Shotkin.

The programs are written in BAL and the operation system is OS/360 (MVT or MFT). The minimum machine configuration required is 256K System/360, a card reader, a printer, four 9-track tape drives and one direct access storage device as temporary storage.

The outstanding feature of this system is its capability to handle the information in its natural free-text form.

The original document is either entered in full, or the text is abstracted and some headings and subheadings (actually the keywords or descriptors or terms or concepts in varying terminology) are picked out to characterize the subject matter. This refers to entering the TEXT-PAC system with one's own data and does not pertain to the use of COMPENDEX tapes where the input is 360 Condensed text 260. This full text is introduced on each punched card by the identification number and print control which provision allows further processing of the information related to the original document and according to various parts of this item (title, citation, author, text, etc.)

The user is offered essentially three types of service (see also Figure 1) originating from the same data base:

1. A Bulletin which lists the transactions to the data base for a given period of time arranged in ascending order of identification number. The key to the Report is the indexes which enable the user to find the information on the basis of category, subject (or subject heading and subheading in COMPENDEX), author. Also KWOC indexes may be produced.

2. Current Information Selection (Selective Dissemination of Information) which keeps the user abreast with the scientific or technical development in his own area. A user's interest profile is matched against the tape containing the transactions of the respective period. The matching documents constitute hits which are disseminated

to the appropriate users.

3. Retrospective Search is a one-time search against a retrospective data base whenever such a need may arise for a particular user. The kind of query submitted to the computer in this case is essentially the same as in CIS, but there is no machine-readable feedback from the user to the system as is in CIS.

In this report we are primarily concerned with Current Information Selection.

4. USERS OF COMPENDEX IN 1969

No.	Surname, Initials	Profession	Profiles	Search Expressions	Words + Symbols	Users U. of C.	Users Outside
1	BROWN, R. A.	Mech.	1	3	35		X
2	JENSEN, E. T.	Mech.	1	3	13		X
3	RACZUK, T. W.	Mech.	2	(1) 6	47		X
				(2) 1	8		
4	WISKEL, A. S.	Chem.	1	1	12		X
5	FITZPATRICK, A.B.	Mech.	1	3	36		X
6	KRUYER, H.S. Ellis	Chem.	1	7	86		X
7	WIGGINS, E. J.	Manag.	1	2	15		X
8	PALLAT, R.	Geol.	1	2	13		X
9	FINLEY, B.	Mech.	1	3	5		X
10	EVANS, I.	Chem.	1	1	1		X
11	ANDERSON, C.	Industr.	1	1	8		X
12	DEBANNE, J. G.	Chem.	1	1	13		X
13	VANDENBERG, A.	Geol.	1	1	23		X
14	ROUND, G.	Chem.	1	1	17		X
15	IMORDE, H.	Mech.	1	1	26		X
16	GAFFNEY, I.	Inf. Retr.	3	(1) 6	31		X
				(2) 5			
				(3) 10	48		
17	GREGORY, J.	Industr.	1	1	12		X
18	VOSS, W. A.	Chem.	1	8	60		X

Users of COMPENDEX in 1969 (continued)

No.	Surname, Initials	Profession	Profiles	Search Expressions	Words + Symbols	Users U. of C.	Users Outside
19	THOMPSON, G. R.	Chem.	4	(1) 4	26		X
				(2) 4	44		
				(3) 2	4		
				(4) 3	8		
20	FEICK, J.	Chem.	5	(1) 1	11		X
				(2) 2	17		
				(3) 3	12		
				(4) 3	22		
				(5) 1	2		
21	TOMIE, M. J.	Chem.	1	3	22		X
22	ANDRE, H.	Chem.	1	19	63	X	
23	AZIZ, K.	Chem.	1	10	43	X	
24	BENNION, D. W.	Chem.	1	19	101	X	
25	DE KRASINSKI, J.S.	Mech.	1	3	34	X	
26	DOIGE, A. G.	Mech.	1	5	40	X	
27	DONNELLY, J. K.	Chem.	1	11	56	X	
28	EDER, W. E.	Mech.	3	(1) 2	15	X	
				(2) 7	29		
				(3) 4	22		
29	GREGORY, G. A.	Chem.	1	24	106	X	
30	GROVES, T. K.	Mech.	1	9	41	X	
31	HARRISON, D.	Civil	1	24	92	X	
32	HEIDEMANN, R. A.	Chem.	1	18	79	X	
33	KRAYER, J.	Mech.	1	4	21		X
34	MIKULCIK, E. C.	Mech.	1	14	88	X	
35	NORRIE, D. H.	Mech.	2	(1)87	256	X	
				(2)19	65		
36	STANISLAV, J. F.	Chem.	1	3	18	X	
37	VENART, J. E.	Mech.	1	7	31	X	
38	KARIM, G. A.	Mech.	1	11	51	X	
39	de VRIES, G.	Mech.	1	4	20	X	
40	HOPE, G. S.	Elec.	1	16	32	X	

Users of COMPENDEX in 1969 (continued)

No.	Surname, Initials	Profession	Profiles	Search Expressions	Words + Symbols	Users	
						U. of C.	Outside
41	DILGER, W.	Civil	1	13	68	X	
42	GAMBLE, B. R.	Civil	1	11	77	X	
43	ROSS, G. A.	Civil	1	20	98	X	
44	COLDHAM, D. G.	Elec.	1	3	15	X	
45	DENNIS, L. P.	Elec.	1	18	42	X	
46	WONG, S. W.	Chem.	6	(1) 1	26		X
				(2) 1	8		
				(3) 3	13		
				(4) 1	14		
				(5) 3	14		
				(6) 2	11		
47	BOMBARDIERI, C. C.	Mech.	6	(1) 1	9		X
				(2) 2	10		
				(3) 1	14		
				(4) 1	20		
				(5) 1	6		
				(6) 1	15		
	TOTAL	Mech. 17	70	496	2471	23	24
		Chem. 17					
		Civil 4					
		Elec. 3					
		Geol. 2					
		Indust. 2					
		Manag. 1					
		Inf. R. 1					
		USERS 47					

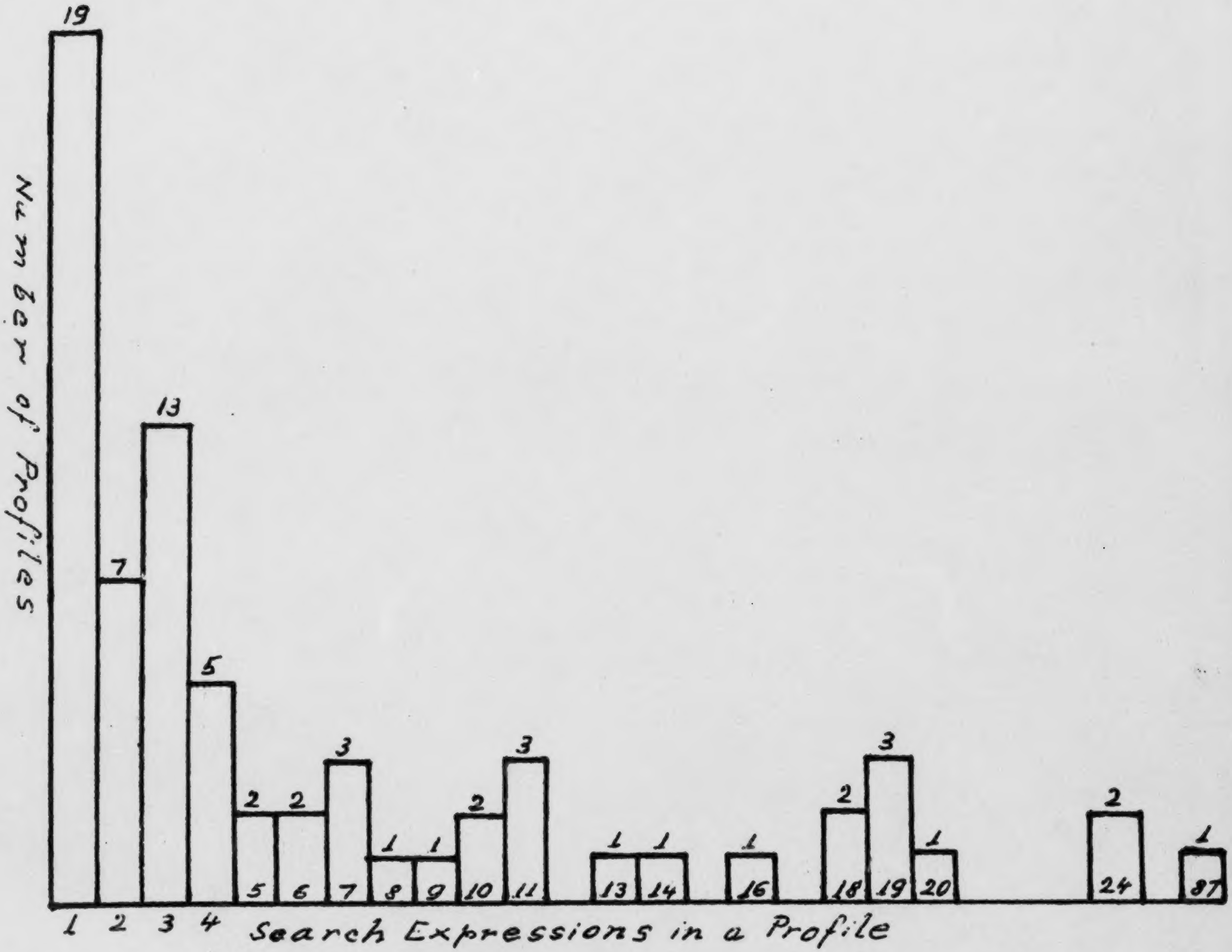


Fig. 2 - Number of Search Expressions per Profile

The users of COMPENDEX system were recruited at the very beginning of our work. The advertising action was taken both on our campus and by AIRA for the Edmonton area. CIS mode was started first and the successful implementing of profile programs was the first task we had to tackle. The decisive factor in the selection of users was their real interest in this work.

The monthly tapes were run in this order and the number of profiles has been steadily increasing:

1969		1970	
January	43	January	75
August	43	February	81
September	43	March	75
July	57	April	82
February	70	May	75
October	70	June	106
November	70	July	106
December	70		

Fig. 3 Number of Profiles Processed

The order of processing the tapes was determined not only by the availability of tapes, but also we wanted to check if the errors in format were present in all tapes throughout the year.

The remaining months of March, April, May, June, will not be processed in the CIS mode, but will be included in the retrospective data base. The reason is that the pilot project is accomplished and running these months would not offer any current information now. The relevant information will be found in retrospective searches for those users who will order a retrospective search.

In July, 1970 the number of profiles processed reached 106.

In 1969, in the total number of users (47) who have submitted 70 profiles, there are 23 from The University of Calgary and 24 from AIRA, Edmonton. The number of profiles per user, search expressions per profile, words per search expression, words per profile, words per user, (average, maximum, minimum) are shown in the table below:

	Average	Minimum	Maximum
Profiles/user	1.5	1	6
Search expressions/profile	7.1	1	87
Words/search expression	5	-	-
Words/profile	35	-	-
Words/user	53	-	-

Fig. 4 Profiles, Search Expressions, Words

Among our 47 users (1969) are the same number (17) mechanical and chemical engineers, four civil engineers, three electrical engineers, two geologists, two industrial engineers, one manager, and one information specialist.

These 47 users have submitted altogether 70 profiles, so that the average number of profiles per user is 1.5, ranging from 1 to 6 maximum. Most of our users (39 i.e. 83 per cent) have only one interest-profile.

Most of these submitted profiles contain a low number of search expressions: 39 profiles from the total of 70 profiles contain 1-3 search expressions, although one non-typical profile contains as many as 87 search expressions. The average number of search expressions per profile is 7.1.

The basic unit of any profile is a word. There are on average 5 words in a search expression, 35 words in a profile, and 53 words per user. When counting the words we considered a word not only natural words but also symbols (A_1, A_2, etc.). It must be remembered that the search time per word may vary depending upon the logic connector used and the number of logic levels (maximum three logic levels allowed).

5. INTERACTION SYSTEM-USERS

This section covers the following topics:

1. Announcing of the service and introducing it to each user on an individual basis.

2. The process of creation of interest profiles of those who decided to subscribe to the service.

3. Optionally screening the output to enhance the precision.
4. The dissemination of the information retrieved.
5. Feedback.
6. Modifying the profiles in close cooperation.

An interesting question is, "What kind of contact with the users is optimal to attain the goal?" There is no explicit answer to this question. Contact in person is to be preferred in announcing the service and advertising it. But stating the interest in narrative form, adding the profile words, their synonyms, antonyms, related terms, exclusions, as well as grouping these terms in logical groups is the responsibility of the user and no one can replace him and do this work on his behalf. Any interference with this responsibility of the user, which is most likely to occur in contacting the user at this stage is harmful and is to be avoided. Other contacts on this interface user/system will be in writing, by telephone or in person if necessary and feasible. Contacts in person become, of course, impracticable with the growth of the number of users.

We decided to run the CIS (Current Information Selection) in the first place, with just as many profiles as to allow us to test the system of current awareness (CIS).

The number of profiles has meanwhile increased to 70. The difficulties due to changes of the abstract format (namely missing last characters on some of the printed lines) were gradually overcome. The profiles were established and adjusted with some users according to their performance in the actual runs. We cooperated closely with AIRA Edmonton in training a search editor and in compiling a basic Users' Manual.

The interaction between system and users has proven, as expected, to be crucial for successfully running this service. We have designed a simple form and a brief introductory letter for the users; we contacted them in person and provided an explanation of some details.

It may well be expected that the user will be more engaged in the searching operation once he has access to an on-line (real time) system enabling him to play a more active part in the game and use heuristic methods of searching the files much similar to the browsing through the library. He will lose some time in searching but definitely gain some time in rejecting irrelevant information. But until such systems are available for routine use, only a precise and detailed statement of user's information needs may eliminate most of the failures in information systems performance and an interface is needed between the system and the users. This should be a continuing, not one-time, cooperation which is made a lot easier for the user now, after the introduction of the double-response cards, new Profile Submission Form, and especially with the COMPENDEX Profiling Guide at hand.

These double cards consist of two halves which are both the same size. The user reads the abstract on the left-hand side, pushes the appropriate box on the right-hand side which is the port-a-punch response card. These response cards indicating the users' attitude toward the information (relevant, irrelevant, document wanted, document not wanted) are the feedback from the user to the system enabling us to correct the profiles when needed and improve the precision and/or recall. The evaluation of these feedback response cards will be done by a special program. There was an important improvement made in the print program: to print the source identification (i.e. title and citation) on the response cards. This saves hours of manual work associated with ordering documents wanted by the users. The purpose of the double cards is threefold:

1. to provide feedback from users,
2. to provide for an easy evaluation of this feedback, and
3. to allow the user to order the document wanted by simply pushing the appropriate box in the response card.

As to how many profiles we can handle on COMPENDEX, there is no mechanically imposed limit on the size of the profiles file, but the limiting factors are:

a. search time economics, and
b. work involved in interfacing the user with the system.

A system running 3,500 profiles per week is known.

The amount of work on the part of the search editor (information specialist, information officer) depends largely on:

1. The number of profiles,
2. The complexity of the profiles (number of logic levels, words, search expressions),
3. The degree of sophistication of the search logic,
4. The willingness and ability of users to cooperate,
5. The experience of the search editor, his tools and state of organization.
6. The stage of implementation being considered (greater in the start-up period),
7. The amount of screening required on computer determined hits,
8. The amount of clerical work the search editor must do.

The steps in profile preparation are:

a. Preparing narrative statement
b. Stating profile words
c. Adding the synonyms, antonyms, related terms, exclusions
d. Grouping the above terms in logical groups (AND,OR)
e. Specifying the connectors and other searching tools (e.g. matching criteria, masking, capitalization)
f. Coding profiles
g. Keypunching profiles

The user may go as far as he willing and able. If the user prepares the profile form in a proper way (as far as step d.) then the search editor can handle up to four profiles a day performing the steps e. and f. only. If he has to replace the user in any of the previous steps, no good result and effectiveness in terms of time and quality may be guaranteed.

It follows from what has been said that the capacity of one search editor is a rather involved problem and for the answer to be fair is necessary to define the terms shown above for each particular case. There is a difference between setting up a profile on the one hand and maintaining it, on the other hand. But it should be emphasized that reworking a wrong profile may be a more tedious work than establishing a new one.

The number of profiles a search editor can handle is reported in one paper to be 20 (with exacting service to the user including screening out hits). Other sources indicate that a search editor can cope with several hundred profiles. COMPENDEX logic is relatively complex thus it seems reasonable that in actual practice, one search editor could maintain some 200 profiles in a favourable environment.

6. THE MONTHLY CIS RUNS (1969)

On the whole, eight monthly tapes (January, February, July, August, September, October, November, December) were processed in this 1969 COMPENDEX pilot project.

Details regarding step times of the programs executed and other particulars may be seen in Figure 5 whereas other characteristics are reflected in Figure 6.

The COMPENDEX monthly tapes 1969 did not contain all the abstracts included in Engineering Index Monthly because of input troubles on the part of E.I. The number of abstracts extended over a range from 1230 to 4848 (average 2785).

Number of hits ranged between 1138 to a maximum of 6301 with an average of 3007.

The ratio Hits/Abstracts has risen until the maximum 1.30 in the last 1969 run, as a result of increased number of profiles. This ratio illustrates how the tape is being utilized to give useful results.

The monthly run will follow on a regular schedule as soon as we obtain the tapes as promised, i.e. the tapes are supposed to be dispatched to us on every twelfth workday of the month.

Only eight monthly tapes were processed and the remaining tapes were added to the retrospective-search data base. In the initial stages of our work we encountered serious troubles with missing last characters on some of the printed lines. As we ascertained later these errors were caused by changing the format of the input on the part of Engineering Index. These errors were eliminated thanks to joint efforts of our group and Dr. Kaufman, the author of the IBM's TEXT-PAC. Some of the

Month	No. of Abstracts	No. of Profiles	No. of Hits	No Hits	Profile Update 001	Profile Diag-nost 002	Profile Print 003	Condens Text Edit 260	Edit Convert 210	Edit Print 203	Memory Load 010	CIS Search 011	CIS Answer Inver-sion 012	Disk Load 013	CIS Print 014
1969															
January	1,642	43	1,352	11	0.04	0.12	0.08	12.86	9.03	0.73	0.56	10.42	0.06	0.45	0.62
February	1,527	70	1,856	11	0.04	0.05	0.09	14.69	8.52	0.72	0.45	8.98	0.06	0.45	0.60
July	2,124	57	1,692	11	0.04	0.12	0.08	20.26	12.07	1.08	0.50	12.54	0.06	0.54	0.40
August	3,738	43	2,673	11	0.04	0.12	0.08	30.40	20.78	1.77	0.49	24.68	0.10	0.95	1.81
September	1,230	43	1,138	14	0.04	0.12	0.08	10.64	7.29	0.63	0.46	8.32	0.05	0.32	0.84
October	3,673	70	4,659	11	0.04	0.11	0.08	32.58	19.61	1.69	0.43	22.95	0.16	1.04	1.71
November	3,631	70	4,387	11	0.04	0.11	0.08	31.59	19.51	1.68	0.47	22.46	0.13	1.09	1.67
December	4,848	70	6,301	8	0.04	0.11	0.08	41.21	25.53	2.33	0.44	28.16	0.19	1.35	1.61

Fig. 5 Step Times of Programs

MONTH INDICATOR	Jan.	Feb.	July	Aug.	Sept.	Oct.	Nov.	Dec.	Total	Average
Number of Abstracts	1642	1527	2124	3738	1230	3673	*3500	4848	22,282	2785
Number of Profiles	43	70	57	43	43	70	70	70	466	58
Number of Hits	1352	1856	1692	2673	1138	4659	4387	6301	24,058	3007
Number of Profiles with no hits	11	11	11	11	14	11	11	8	88	11
Ratio Hits/ Abstracts	0.82	1.22	0.80	0.75	0.81	1.27	1.25	1.30	-	-

Fig. 6 Monthly Runs

*Estimate

abstracts were mutilated and we were promised to get an additional tape with this missing information.

In a random sample of 1,000 abstracts we have found 72 misspellings. It is necessary to go on checking these misspellings, as, in full text processing, they could cause some relevant abstracts to be missed.

7. PERFORMANCE OF THE SERVICE

The determination of overall effectiveness of any information system is a very complex problem and the appraisal may be approached from different viewpoints. The ultimate criterion is user satisfaction. The user will consider:

1. The time span between his order and the delivery of the information desired.
2. The cost of the information.
3. The effort needed on his part to get the information (ease of accessibility). In this context he highly appreciates a good relevance.
4. The promptness with which the original (or copy of) information may be obtained if any references (with or without abstracts) are delivered.
5. The appropriateness of the data base to his information need. Related to this is the capability of the processing system to retrieve the desired information.
6. The timeliness of the information contained in the data base.
7. The accuracy and reliability of this information (the quality of indexors' work and of the source).
8. The source language (translation required).

The user should examine all these questions carefully before he subscribes to any information service.

Good rating in these eight points is a prerequisite for any information system to be acceptable for a particular user. If the system fulfills the expectations of the users, then it really has good effectiveness--the effectiveness being the ability of the system to do the job for which it was primarily designed.

In the current practice which is reflected in the literature, several measures of system performance are used and defined. No one is generally accepted and all of them are subject to strong criticism. In the following we will attempt to utilize some of them outlining their merits and demerits.

7.1 Relevance

Let us first consider the relevance called also precision (ratio) or interest ratio. Relevance is the proportion of retrieved relevant documents to all documents retrieved, both relevant and irrelevant. This relevance may lie, as reported in the current literature, anywhere between 18 and up to over 80 per cent.

Relevance is usually judged on the basis of the users' feedback in some form or other. The first problem here is to get the feedback from enough users to allow us to make some valid conclusions. Whereas some workers have received feedback from 80 per cent of their users, others had to put up with considerably less--about 50 per cent.

RELEVANCE ASSESSMENT
(Per cent)

No.	Users	Jan.	July	Aug.	Sept.	Dec.	Note
1	R. A. Brown	-	-	-	-	32	
2	E. T. Jensen	-	-	-	-	0	
3	T. W. Raczuk (000003)	-	-	-	-	100	
4	T. W. Raczuk (000004)	-	-	-	-	-	
5	A. S. Wiskel	-	-	-	-	-	
6	A. B. Fitzpatrick	-	-	-	-	29	
7	H. S. Ellis Kruyer	-	-	-	-	71	
8	E. J. Wiggins	-	-	-	-	-	
9	R. Pallat	-	-	-	-	-	
10	B. Finley	-	-	-	-	63	

Relevance Assessment (continued)

No.	Users	Jan.	July	Aug.	Sept.	Dec.	Note
11	I. Evans	-	-	-	-	10	
12	C. Anderson	-	-	-	-	-	
13	J. G. Debanne	-	-	-	-	-	
14	A. Vandenberg	-	-	-	-	-	
15	G. Round	-	-	-	-	9	
16	H. Imorde	-	-	-	-	-	
17	I. Gaffney (000017)	-	-	-	-	-	
18	I. Gaffney (000018)	-	-	-	-	45	
19	I. Gaffney (000019)	-	-	-	-	-	
20	J. Gregory	-	-	-	-	-	
21	W. A. Voss	-	-	-	-	89	
22	G. R. Thompson (020001)	-	-	-	-	11	
23	G. R. Thompson (020002)	-	-	-	-	-	
24	G. R. Thompson	-	-	-	-	-	
25	G. R. Thompson	-	-	-	-	-	
26	J. Feick (020005)	-	-	-	-	-	
27	J. Feick (020006)	-	-	-	-	-	
28	J. Feick (020007)	-	-	-	-	-	
29	J. Feick (020008)	-	-	-	-	-	
30	M. J. Tomie	-	-	-	-	-	
31	J. Feick (020010)	-	-	-	-	20	
32	J. Krayer	-	-	-	-	-	
	AVERAGE	-	-	-	-	40	AIRA
33	H. Andre	64	77	57	59	77	
34	K. Aziz	38	71	*	*	59	
35	D. W. Bennion	30	18	15	46	34	

Relevance Assessment (continued)

No.	Users	Jan.	July	Aug.	Sept.	Dec.	Note
36	J. S. de Krasinski	*	*	*	*	*	Serv. dis.
37	A. G. Doige	*	*	*	*	*	Serv. dis.
38	J. K. Donnelly	*	*	*	*	15	
39	W. E. Eder (100007)	*	*	*	*	74	
40	G. A. Gregory	*	*	*	*	67	
41	T. K. Groves	*	*	*	*	61	
42	D. Harrison	*	*	*	*	48	
43	R. A. Heidemann	*	*	*	*	85	
44	E. C. Mikulcik	33	39	44	27	43	
45	D. H. Norrie (100014)	50	33	100	0	0	
46	D. H. Norrie (100015)	44	60	24	30	25	
47	J. F. Stanislav	*	*	*	*	50	
48	J. E. Venart	86	*	*	*	68	
49	G. A. Karim	*	*	*	*	100	
50	G. de Vries	75	33	45	40	21	
51	G. S. Hope	13	*	*	*	63	
52	W. Dilger	*	*	*	*	88	
53	B. R. Gamble	56	33	30	31	95	
54	G. A. Ross	*	*	*	*	62	
55	D. B. Coldham	*	*	*	*	100	
56	L. P. Dennis	*	*	*	*	*	Serv. dis.
57	W. E. Eder (100026)	*	*	*	*	52	
58	W. E. Eder (100027)	*	*	*	*	82	
	AVERAGE	44				60	The Univ. of Calgary
59--70	S. W. Wong and C. C. Bombardieri	These users only tried their profiles.					

There were 26 profiles in the Section 2 (Calgary). We distributed 104 answers to them covering the months of January, July, August, September, and we have received 32 responses (31 per cent). In December we received 23 responses from 26 profiles (88 per cent). Obviously this increased response from the users was due to the improved form of the output on double response cards. This form made the evaluation a lot easier both for the user and ourselves. The form of feedback (its convenience) determines very clearly the quality and quantity of the feedback retrieved (its completeness and timeliness).

The average relevance in December was 60 per cent as compared to the average of the previous month's 44 per cent; it indicates a better quality of profiles.

The information may be judged as to whether it is or is not relevant, by the user, by the information specialist or by a jury, which is more objective but is hardly practical. We expect the user to do this. Initially, we supplied the user with hits as presented by the system, without previously scanning them. In 1970 we began to pre-scan the hits and this proved to be effective in enhancing the relevance.

In order to assess the relevance of the information we use the double cards which consist of the abstract on the left-hand side and of the response card at the right-hand side. This response card bears the card number which is also repunched, and gives the instructions how to properly handle it. The user reads the abstracts and makes the judgement of the relevance by pushing out the appropriate box of the port-a-punch card by means of a sharp pencil.

He has the following choice:

Abstract relevant
Abstract irrelevant
Document wanted
Document not wanted
Comments, questions, address change (use reverse side).

If the document is relevant the user has to push out two boxes denoting "relevant" and either "document wanted" or "document not wanted."

In the experimental stage these response cards are manually processed but provision is made to do this automatically by a computer program.

Relevance Per Cent	Number of Profiles
100	2
90-100	1
80-90	3
70-80	2
60-70	5
50-60	3
40-50	2
30-40	1
20-30	2
10-20	1
0-10	1

Fig. 7 Relevance of Output (1969)

This table (Figure 7) gives a picture that is in good agreement with the average value as it indicates the highest number (5) of profiles in the vicinity of 60 per cent. Both extremes (0 and twice 100) are non-typical.

Feedback and relevance for Calgary and Edmonton are represented in the tables following (Figure 8):

DATA FOR 26 PROFILES (SECTION CALGARY)

Period (1969)	Jan., July, Aug., Sept.	December
Feedback received		
Users	8 (Average)	23 (of 26)
Per cent	31	88
Relevance (per cent)	44	60

DATA FOR 32 PROFILES (SECTION AIRA)

Period (1969)	Jan., July, Aug., Sept.	December
Feedback received		
Users	-	12 (of 32)
Per cent	-	38
Relevance (per cent)	?	40

Fig. 8 Feedback Received and Relevance (1969)

In 1970 we have been receiving feedback in some form or other concerning 92 per cent of profiles (23 of 25). Relevance in the first seven months has been 76, 73, 69, 47, 54, 55, 68 per cent.

The relevance as a measure of information system effectiveness is widely used. The main objection against it is that it is based on the subjective judgement of the user. It might be said that it is "a precise calculation of inaccurate data." If the performance of the system is to be appraised, then there must be a complete coincidence between the information need and between the interest profile of the user. Otherwise there is a distinct discrepancy between the relevance seen through the interest profile and that seen through information need, for the same information supplied. The judgement of the same user may vary depending on what stage of work he is currently engaged in. In addition to this time dependance there is also a place dependance which plays a part in the judgement if a particular information is or is not relevant:

1. the source of information is out of reach within a reasonable period of time,

2. the idea described is not practicable locally.

In all these cases the user should be instructed to denote such an information as "relevant--document not wanted," (if such facility is built into the feedback response) rather than "irrelevant."

Apparently, we are involved here in psychological aspects of information retrieval which area was not yet explored at all. We have found that users' judgement as to whether an information is or is not relevant, may be influenced also by the fact that the user has got some information which he considers to be a big hit and any other information is overshadowed by this previous one and is more likely to be estimated "irrelevant." Sometimes the information need of the user is satisfied at a certain point and further information is of no interest tending to be marked "irrelevant"; this may happen if the user is looking only for some ideas or inspiration and such a user is very fastidious. The reverse is true with a user who needs a complete, exhaustive search covering a special area of interest, e.g., a patent search opening a research project. Such a user wants to see many documents to make sure he does not duplicate the work that has already been done elsewhere and/or that he does not infringe other people's rights. Such a user tends to denote the information rather as "relevant" "document wanted."

Also the user tends to mark the information as "irrelevant" if he has seen it before which is, of course, incorrect. If he considers the content to be of poor quality, he might also mark "irrelevant."

It should be emphasized at this point that user's appraisal of the information supplied is much easier in full text processing services than in services giving the title, author, and citation without any text. Such services leave much to the user's imagination to decide if the information pertains to his interest. This may shift the relevancy figure up or down but always at the expense of accuracy.

Perhaps the most interesting is that the users sometimes label irrelevant information as relevant, if it brings some inspiration outside the profile.

When evaluating the relevance we must not forget that this is no absolute measure but rather an imperfect tool for estimating the performance of the profile in a given environment of the system, data base, computer, user, and search editor. The practical point here is whether or not the user himself is satisfied. Some users are content with a relatively low per cent-relevance, whereas others are unsatisfied with a considerably higher relevance. Generally a user tends to judge the service more favourably if he gets ten items two of which are relevant, than if he gets 150 items, thirty of which are relevant.

It is one of the paradoxes in this field, that most users highly appreciate if they are not inundated by irrelevant information even if they are unknowingly losing much of the information which could have been retrieved had the search been conducted at another relevance: recall trade-off.

An interesting point in this context is to compare, (1) a system searching the keywords (concepts, terms, descriptors) assigned to documents, (2) system with searching based on titles, and (3) a full text processing system, although this topic goes a little beyond the objective of this section. We will also use the term "recall" which will be dealt with next. Let us use the terms "exhaustivity" and "specificity" accepted by the Cranfield Project and coined by F. W. Lancaster (Information Retrieval Systems, Characteristics, Testing and Evaluation; 1968, John Wiley & Sons Inc.) which made a valuable contribution both to theory and practice of retrieval systems evaluation.

In order to understand the problem of relevance in its full significance we must examine two sets of descriptions:

A. Description of documents
   1. keywords in the system
   2. title in the system
   3. full text (mostly an abstract) in the system

B. Description of user's interest
    1. keywords in the system
    2. profile in the system
    3. profile in the system

We know that any hit is produced as a result of a match between A (description of documents) and B (description of user's interest).

Description of documents (A) may be, as far as relevance is concerned, more or less exhaustive (i.e. contain more or fewer expressions pertaining to different categories or facets) and more or less specific (finely defined, higher on the hierarchy tree). Exhaustive A means higher recall and may entail lower relevance; a specific A implies higher relevancy and may cause reduced recall. The specificity and exhaustivity in the system (1) reflects the responsibility and capability of the indexor and/or the indexing policy adopted. The specificity and exhaustivity of the title (2) is in many cases rather limited. The full text processing (3) has definitely the good chance to offer both a fair exhaustivity and specificity provided an expert abstracting work has been done. The professional abstractor must have due regard to all the categories (facets) describing the subject matter, as well as to various degrees of specificity, leaving out all the unnecessary ballast which claims the costly storage and increases the cost of computer processing.

Only such a data base enables us to search in a wide range of recall and relevance values at the discretion of the search editor. The foundations for a well-balanced and meaningful search are laid right here. It should be noted that even the best formulated profile or question will not find a satisfactory answer if the data base is not properly constructed. This is of special significance in systems with highly sophisticated searching capabilities which would be all in vain with a data base not allowing their full utilization.

In addition to exhaustivity and specificity there is another dimension which plays an important part both in the data base and the query: we may call it "synonymity." It means how completely synonyms (and antonyms and related terms, if applicable) are specified. Synonymity is characterized by "OR" in queries.

The role of exhaustivity, specificity, and synonymity, both in the data base and query, towards the relevance and recall may be visualized by the table below (Figure 9):

Where applied / Dimensions	Data Base	Query (profile, question)
(high) exhaustivity	(high) recall	(high) relevance [1]
(high) specificity	(high) relevance	(high) relevance [2]
(high) synonymity	(high) recall	(high) recall

[1] High relevance will result if we apply high exhaustivity within the search expressions. If we, however, apply the exhaustivity by using more search expressions (multiple approach), this will entail an improved recall.

[2] If we do not want the recall to be impaired, we have to use as many hierarchical levels as needed, i.e., various degrees of specificity connected by OR.

Fig. 9 Dimensions in Indexing and Query Formulation

The following figure suggests a three-dimensional framework for representation of a document and/or query description (Figure 10). Together with the table above, it shows how to use these dimensions to monitor the output in the direction desired.

Descriptions of user's interest-query (profile or question) are characterized by a certain degree of the same dimensions as was the data base. However, they do not necessarily influence the result of a query in the same way as if they were applied in the data base (see Figure 9). It is obvious that both high exhaustivity and specificity will tend to enhance the relevance and reduce the recall. Such one-sided improving the relevance is mostly regarded as a detrimental phenomenon in the retrieval system's performance. The recall may be improved by incorporating higher degree of synonymity to the query. See also notes 1 and 2.

The synonymity (specifying synonyms), of course, is not too

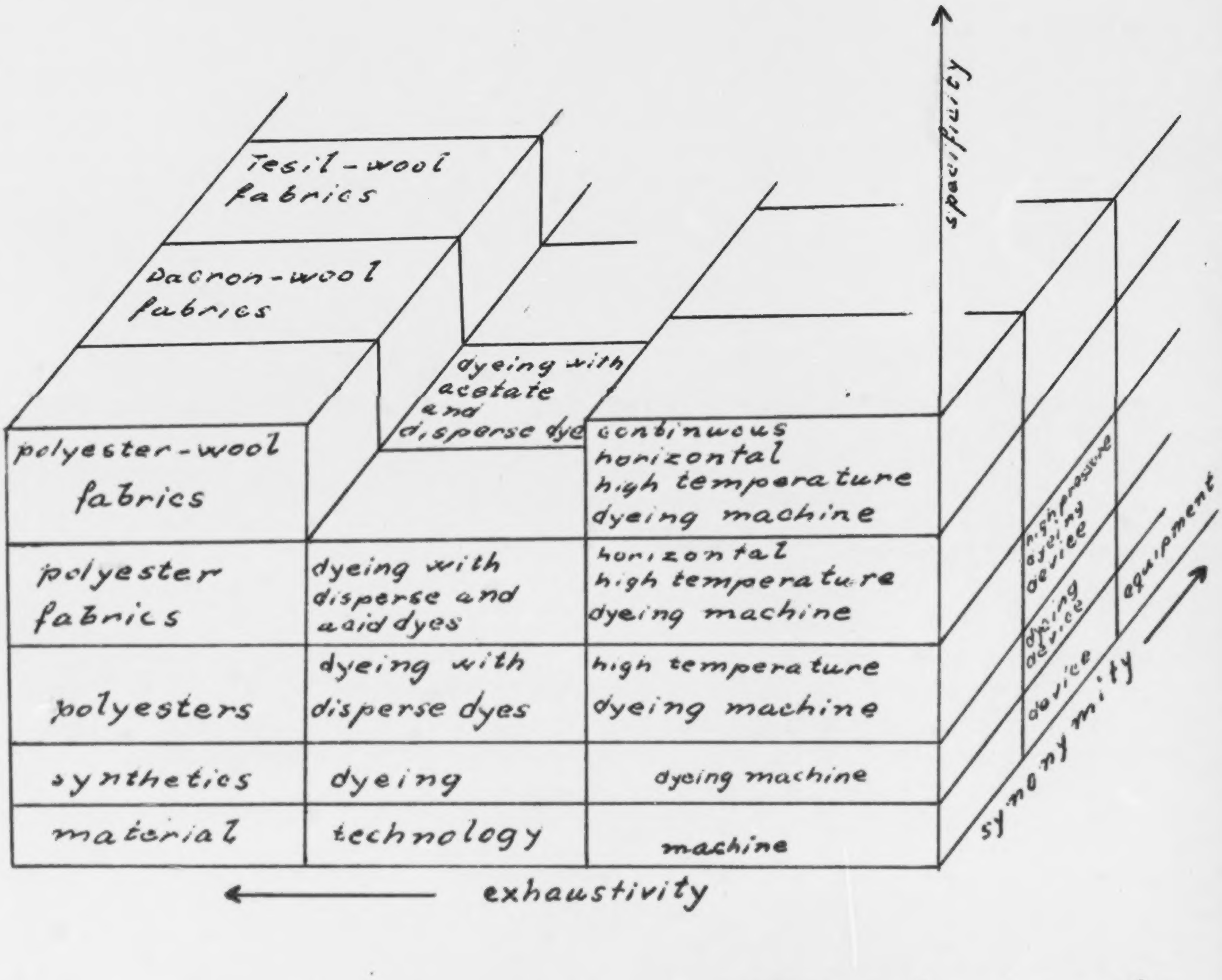

specificity
Tesil-wool fabrics
Dacron-wool fabrics
polyester-wool fabrics
polyester fabrics
polyesters
synthetics
material
dyeing with acetate and disperse dyes
dyeing with disperse and acid dyes
dyeing with disperse dyes
dyeing
technology
continuous horizontal high temperature dyeing machine
horizontal high temperature dyeing machine
high temperature dyeing machine
dyeing machine
machine
high pressure dyeing device
dyeing device
device
equipment
synonymity
exhaustivity

Fig. 10 - Exhaustivity, Specificity, Synonymity

significant:

1. when a controlled vocabulary is used both for indexing and search formula establishing (indexing systems),

2. when a dictionary is automatically generated listing all words occurring in the data base, which enables the search editor to set up the profile (question) accordingly.

One example will elucidate these principles. The user needs information on the topic "machine for the dyeing of synthetic fibres." We want to question a data base which is supposed to contain abstracts oriented to this subject matter.

Our terms (words) are "machine," "dyeing," and "synthetic fibres" (Figure 10). It is evident that an exhaustive formula covering all of these terms (taking into account the facets equipment, technology, and material, represented by these three terms) will bring about a high relevance. Our tools in the TEXT-PAC system by means of which we may connect these three terms are "AND", "WITH", "ADJACENT" and they offer us a very desirable additional capability to control the recall (see Figure 11). Obviously, the highest recall will result from the connector "AND", lower recall will result with "WITH", and practically no answer (in this particular case) will be received with "ADJACENT". "ADJACENT" is used to increase the relevance. It makes the profile or question more specific and may be used only if the words of the expression occur close to each other, otherwise it endangers the recall. The third way of governing the relevance and recall is by including synonyms, antonyms, and related terms into the search formula. If we use the synonyms "chemical fibres" and "artificial fibres" in addition to "synthetic fibres" in the query, we improve the recall without deteriorating relevance. If we use "polyamide fibres" instead of "synthetic fibres" as defining more precisely our special interest, in other words if we proceed in the direction towards a higher specificity, we increase the relevance and may adversely affect the recall. The synonyms and antonyms are, of course connected by the Boolean "OR". (Regarding Dictionary see above.)

Logical Connectors	Function	
	Recall	Relevance
AND	↑	↓ (If Logical Connectors Applicable)
WITH		
ADJACENT		

Fig. 11 Control of Relevance/Recall by AND, WITH, ADJACENT

The TEXT-PAC system and some other systems have additional means of how to monitor the output. The masking (truncation) will promote, like the synonyms, the recall and may, if not properly stated, affect the relevance. Relatively seldom used is the "CONTROL" which restrains the search only to one or more print controls and, therefore, yields a limited output with a lower recall without improving the relevance. For example, we may, for any reason whatsoever, restrain the search to the titles exclusively and we miss all matches in other print controls (worse recall), but we have not guaranteed better relevance, because the searching logic remains the same. The operator "NOT CONTROL" has a much similar effect. The use of higher match criterion has also a restrictive effect on the output with a lower recall; in this case, however, the relevance may be fostered if the concepts matched are related to the same subject being searched.

It should be noted that the TEXT-PAC system creates automatically a very useful tool for the search editor: the dictionary of words occurring in the data base. Although the generation of this dictionary involves additional computer time, it is invaluable in setting up profiles as it ensures that the same vocabulary be used in profiles as was in the data base. Using this dictionary we may improve the overall performance of the system.

It is an inherent property of search formulation in TEXT-PAC that any concept may be constructed with three levels of logic structure. It is apparent that using these "vertical structures," as we would like to call it, we aim to a higher specificity and/or exhaustivity and we attain a better degree of relevance. The following example (Figure 12) is designed to demonstrate what we have meant under "three levels" and "vertical structure":

Grade of Logic Level	Logic Symbol	Words or Logic Symbols
0	A1	Information ADJ Retrieval
0	A2	Comput $$$
1	A3	A1 AND A2
0	A4	Canada
0	A5	USA
0	A6	United ADJ States
0	A7	United ADJ States ADJ or ADJ America
0	A8	North ADJ America
0	A9	North-America
1	A10	A5 OR A6 OR A7 OR A8 OR A9
2	A11	A3 AND A10
0	A12	Universit$$$ OR Campus$$ OR College$ OR Educom
3	CON 1	A11 AND A12

Fig. 12 Levels of Vertical Structure

From what has been said it may be concluded that there is a pronounced trade-off between relevance and recall. Recall is not considered in the evaluation of many systems and this is due to either the elaborate methods used to assess it or because of mistrust of methods based on statistical samples.

There are some other methods available on how to evaluate the relevance. One of them does not take into account all of the relevant abstracts but only those ones which are regarded worthwhile to order a copy or original of the document. In our opinion this method represents no refinement but aggravates the evaluation by additional inaccuracy: maybe the user himself or his staff procures the copies or the copies will be ordered later when needed, or the user studies the original source in the library.

A much more reasonable approach to estimating the success or failure of the service seems to be to estimate what is the proportion of our cards among the information items which the user considers to be most significant. But this method involves two subjective judgements: what is most significant and what is the proportion of our cards. Accordingly the accuracy of this approach represents no progress.

7.2 Analysis of Relevance

Regarding relevance (precision) it is common and useful to establish the relevance figures. They are some indication of the user's satisfaction, especially over a certain period of time. They can be a warning that something is wrong in serving a particular user. We must be very careful when comparing individual users or user groups. Comparing various systems by means of relevance values requires a thorough consideration of many factors (users' judgement, relevance/recall preference, method of calculating the relevance - ratio of averages versus average of ratios, user/system interface, logic tools, etc.).

Even more meaningful than to calculate the relevance figures is to examine the relevance failures. This means to find out why a certain abstract was selected which, later on, was rejected by the user as irrelevant. The reasons for failures should be sorted into groups and expressed in terms of percentage. This analysis should enable us to adopt efficient steps to avoid failures as far as possible. We should be, however, fully aware of what we want to achieve for any particular user in terms of relevance/recall trade-offs. In other words, some sound compromise must be found which appears to be the most acceptable to the user.

(A similar procedure is applied to the recall failures.)

In our assessment, analysis, and results evaluation we have used users' feedback cards indicating "irrelevant" abstracts. We were tracing the failures for the months of January, February and March, 1970. Our investigation was limited to the users who forwarded their response (feedback) cards to us in due time. Altogether one hundred failures were examined.

Theoretically failures may be divided into the following groups indicating their causes:

0. Users

Users denote some abstracts as irrelevant although they really match the profile. This is not a failure of the system at all. The user simply rejects information to which he assigns a minor or no value.

1. Abstracts

If words were used in the abstract which do not properly describe the content, then the abstract found will be irrelevant. This irrelevance may sometimes come out only after delivery of the hard copy. It is a failure of the abstractor not of the retrieval subsystem.

2. Questions

2.1 If the terms used are not appropriate, irrelevant abstracts will be retrieved (see also recall).

2.2 If terms used are not sufficiently specific non-pertinent information might result (a bad recall in the reverse case). In this case the question is broader than the user's need.

2.3 If the question (any one search expression) is not exhaustive enough (also in a restrictive sense) the relevance could be impaired (a bad recall in the reverse case).

2.4 Improper search logic may affect the relevance, producing irrelevant output. This implies incorrect use of logical connectors, truncation, incorrect set-up of search expressions from the concepts, etc.

2.5 Ambiguous terms also deteriorate relevance. Different authors with identical names, words occurring in journal titles, homonyms, belong in this subgroup.

2.6 Although the question is well formulated, some abstracts are found to be irrelevant due to a false coordination. (A false coordination may result also under conditions given e.g., under 2.3, 2.4 and 2.5).

3. Computer, programs

These are other possible sources of relevance failure.

4. Coding, typing, punching could also produce some irrelevant information.

The following table (Figure 13) illustrates which percentage of relevance failures is to be attributed to the groups indicated above.

Group	0	1	2.1	2.2	2.3	2.4	2.5	2.6	3	4	Total
Per Cent	12	0	0	6	53	4	3	12	9	1	100

Fig. 13 Relevance Failures

We may conclude from the figures shown:

0. Users should be instructed once again about the meaning of "relevant" and "irrelevant." "Irrelevant" by no means should be used to denote the information which is pertinent to the profile as it was specified. If the user has a negative attitude to such an information, it should be labeled as "relevant, not wanted." If the information need has changed in the meantime, the profile should be changed for the feedback to be meaningful.

1. There was no one failure which could be attributed to the quality of abstract. It should be remembered that some of such errors might be discovered only after delivery of the hard copy respective; the retrieval centre is mostly not kept posted by the user of such failures.

2.1 The terms used in the questions have not caused any failure.

2.2 Little specific (too broad) terms were the reason of failure in 6 per cent of all failures examined. There are, of course, certain restraints in moving the specificity up and down in any particular case. This depends on how the user is oriented: relevance-oriented or recall-oriented, or compromise.

2.3 53 per cent of all failures under review goes to the account of little exhaustivity.

Although we have set up separate groups 2.2 and 2.3 for little specificity and exhaustivity respectively, we feel, that in the most cases, it is hard to draw an exact boundary. In many instances both higher specificity or exhaustivity could bring about a better relevance. Both 2.2 and 2.3 are responsible for 59 per cent of failures. Here is the most sensitive tool for monitoring the desired relation between relevance and recall.

2.4 4 per cent of all failures were due to a faulty search logic (truncation - 2 per cent, logical connector - 1 per cent, formulation of search expression using concepts - 1 per cent).

2.5 Ambiguous terms represent 3 per cent. They can be obviated by using more exhaustive formulation.

2.6 There is not much that can be done about this 12 per cent share in failures. Any change either is difficult to make or it would have other hazards to it.

3. Hardware or software is to be blamed in nine cases out of 100 failures.

4. There was only one error in typing, coding, punching responsible for a relevance failure.

Summing up, we can state that the correct formulation of a question is the best guarantee for a good relevance. A defective question was behind 78 per cent of all failures. The share of searching tools (2, 4) was relatively negligible.

It appears that our attention should be focused to the right proportions in the specificity and exhaustivity of concepts and search expressions. This is only possible if we know, for each individual profile, the orientation either to recall or relevance or any compromise. The best solution to this problem seems to be subdividing the users into three categories.

Though our examination was based on 100 relevance failures only, the results are conformable to our daily experience.

We recommend to continue this type of analysis. It is the best indicator of what should be done with any individual profile and with the service as a whole.

7.3 Recall

In estimating the recall of some of the profiles we were aware that we cannot count on the cooperation of the users, because it would take too much of their time. We also realized that it is not feasible to establish the recall values for 70 profiles by the means available, using conventional method of screening out the entire data base. On the other hand, we strongly felt that, unlike some other workers who content themselves with relevance figures only, we need at least some more or less precise recall figures to complement the picture of the system performance as outlined by the relevance figures.

After a careful consideration of the goals to be achieved, the means and time available, we evolved the following method.

This method does not involve all of the documents because of the size of the data base (4848 abstracts, round 5000) under evaluation and the number of profiles (70). The features of this method are:

1. The judgement was done by an information specialist rather than by the user. A careful selection of profiles has made it possible. The profiles were compared against the data base successively. Each time one profile was thoroughly studied as well as the documents which were indicated as relevant by the user.

2. Only samples were taken from the data base rather than scanning the entire data base.

3. Actually, we should have excluded the relevant documents retrieved from our scanning, but we left them deliberately if they happened to be in the random sample taken; we used them as a check that we were proceeding correctly as would most likely the user proceed. If we did not find all the information the user had marked "relevant" (in course of relevance evaluation), this would mean that we have not properly understood the user's information need as expressed in the

profile and that we are unable to estimate the recall figure for this particular profile. We can take the samples in such manner that we always include one or more relevant items to check the consistency of scanning.

4. We do not consider relevant the information which was rejected by the user as irrelevant.

The best method is to determine recall values for high, medium and for low relevance values. These recall values are supposed to be on the lower side as well as on the higher side, respectively. This would enable us to draw a relevance/recall curve. This curve indicates approximately in which region we are operating our system.

Another important consideration is what is the right size of the sample taken.

Let us take the profile number 100018 which has achieved 100 per cent relevance of output in the month of December, 1969. The number of relevant responses was 10. The number of records in the data base was 4848 (or roughly speaking 5000). Theoretically, we should find in a sample of 500 records one relevant abstract.

Minimum size of any sample examined should, therefore, be

$$\text{Smin} = \frac{A}{Rr}$$

where A means number of abstracts in the data base, Rr stands for "Retrieved relevant."

Instead of Smin we can, of course, use any of its multiples, maximum being the entire data base. It depends on which amount of abstracts we consider manageable. The larger the sample, the more reliable results we get. In our example we could use 500, 1000, 1500 and so forth, abstracts.

In our examination of the 500 abstracts (profile 100018) we found three abstracts which could well be considered relevant to the information need specified and were not retrieved in actual run. At the same time we should have found (statistically) one relevant retrieved abstract; this abstract (also none or more than one could be retrieved in manual scanning) is our check that we understand the relevancy for this particular profile.

Finding 3 additional relevant abstracts in 500 abstracts implies that 30 abstracts should be theoretically found in the whole data base. The number of all relevant abstracts, retrieved (10) and not retrieved (30), would be 40 and recall for this profile would be 25 per cent.

In our evaluation method we calculate the recall as

$$\text{Rec} = \frac{E}{E + \text{Relnr}} \times 100$$

where E = number of relevant retrieved abstracts theoretically expected to be in the given sample, Relnr = relevant abstracts not retrieved found in the sample examined.

Recall for the profile 100018 was, therefore,

$$\text{Rec} = \frac{1}{1 + 3} \times 100 = 25 \text{ per cent}$$

If we took the sample of 1000 abstracts (2 x Smin) and if we found Relnr = 6, then

$$\text{Rec} = \frac{2}{2 + 6} \times 100 = 25 \text{ per cent}$$

Although this method cannot be claimed as completely reflecting the virtual recall, no method can. Each of them is encumbered by subjective judgements stating the relevance. But the same applies to it, as to any other method based on statistical premises: it is a useful measure of recall if it is used consistently throughout all the project.

We recommend a continuous analysis of recall failures as one means of keeping the recall values at the level desired for each individual profile.

The following recall values (see Figure 14) were established.

Altogether 6730 records were scanned for eight profiles and sixteen relevant abstracts not retrieved were found in the samples.

This method of recall estimation is suitable for an SDI service. For retrospective searches it would be hardly practical in view of the bulky samples that would be necessarily involved for a large data base (particularly with a small number of relevant retrieved). In this case the method based on retrieving a certain number of relevant documents known to be in the data base might be the only feasible one. It would require cooperation on the part of the users.

No.	Name	Profile	Relevant Retrieved	Sample	Relnr (in sample)	Recall	Formula
A	Coldham, D.B.	100024	5	1000	4	20	$\frac{1}{1+4}$ x 100
B	Karim, G.A.	100018	10	500	3	25	$\frac{1}{1+3}$ x 100
C	Dilger, W.	100021	7	714	2	33	$\frac{1}{1+2}$ x 100
D	Ross, G.A.	100023	16	313	2	33	$\frac{1}{1+2}$ x 100
E	Groves, T.K.	100009	36	2 x 139	1	67	$\frac{2}{2+1}$ x 100
F	Eder, Q.E.	100026	114	4 x 44	2	67	$\frac{4}{4+2}$ x 100
G	Harrison, D.	100010	12	417	1	50	$\frac{1}{1+1}$ x 100
H	De Vries, G.	100019	3	2 x 1666	1	67	$\frac{2}{2+1}$ x 100

Fig. 14 Recall Values for Selected Profiles Output

7.4 Analysis of Recall

Having calculated the recall figures we examined some of the recall failures. In other words, we turned our attention to the "relevant, not retrieved."

Doing this we went through the data base sample and tried to find out why the relevant abstract was not retrieved in the actual run. The reason could be one of the following:

1. Questions

    1.1 Terms used are wrong ones, we may expect a recall failure (and relevance failure at the same time).
    1.2 The terms used are too specific; the same outcome may be expected (the need broader than the question).
    1.3 The question is too exhaustive; the result will be low recall.
    1.4 The question does not include all aspects of the need; the recall will be reduced. Aspects should be covered by separate search expressions to enhance recall, otherwise you increase exhaustivity of a search expression and you promote relevance.
    1.5 Not all synonyms are specified; there will be a decline in recall (this may happen even if you have Word Frequency or Dictionary).
    1.6 Improper logic is used (logical connectors ADJ, WITH where AND would do, incorrect truncation, etc.)

2. Hardware, software failures.

3. Coding, typing, punching failures.

The following table (Figure 15) is indicative of what has caused the recall failures examined.

Type of Failure	1.1	1.2	1.3	1.4	1.5	1.6	2	3	TOTAL
Number of Failures	0	1	3	2	2	8	0	0	16
Per Cent	0	6	19	12.5	12.5	50	0	0	100

Fig. 15 Recall Failures

It may be concluded from these figures that the best recall will be achieved by a proper question formulation. This implies a correct logic (50 per cent) as well as other characteristics of a good question (1.2 through 1.5). The amount of specificity and exhaustivity will act on the balance between relevance and recall.

Although we are operating here with a relatively small number of results, these were gathered by scanning large data base samples and very diversified profiles.

7.5 Precision - Recall

Having established some relevance and recall figures, the next logical step was to investigate how they relate to each other for the given profiles. Figures 16 and 17 illustrate the plotted and tabulated values:

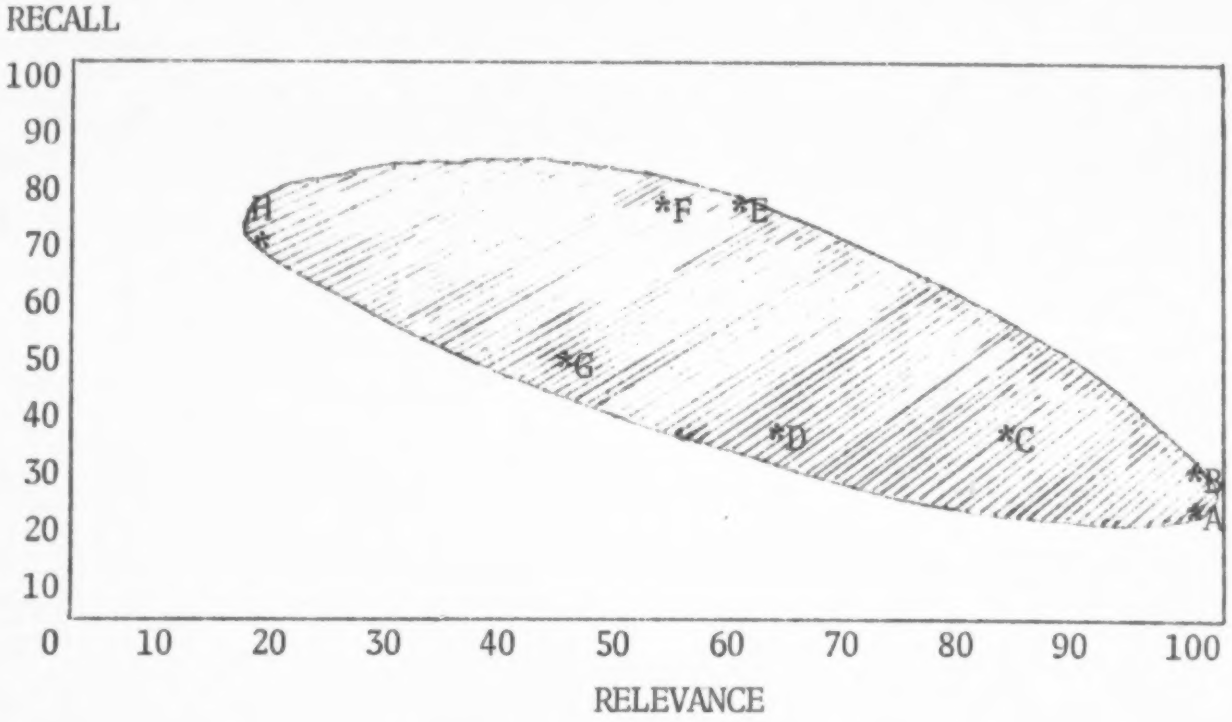


Fig. 16 Relevance/Recall Graph

Profile	Point	Rel	Rec
100024	A	100	20
100018	B	100	25
100021	C	88	33
100023	D	62	33
100009	E	61	67
100026	F	52	67
100010	G	48	50
100019	H	20	67
*Average Per Cent		66	45

Fig. 17 Relevance/Recall Table

We could not draw the curve for all our profiles because of lack of recall figures. However, it may be expected that this plot is roughly representative for all profiles run as we have chosen profiles from the highest to a low relevance. The indirect relationship between relevance and recall was substantiated once more; it is illustrated in the tabulated values as well as in the graph.

This graph demonstrates nothing more and nothing less than the relationship of relevance and recall of eight selected profiles (for which there were recall values available) in the December, 1969 run. It would be very interesting to have plots for:

1. all profiles individually in any monthly run,

2. all profiles individually over a longer period of time (averages).

3. monthly runs as a whole, over a longer period of time (monthly (averages).

From our graph we can see that we are operating in a reasonable region in the middle of the field. This pertains to the system as a whole.

This graph, however, may be used as a measure of satisfaction of individual users. It is clear that a system is only good when it makes the users happy. This means that this particular system is considered good by the user, if users A and B prefer high relevance at the expense of recall, whereas users E and F like some compromise in between. User H is inclined to accept low relevance and favours good recall (which could be further improved).

To insure the satisfaction of the users in the way described it is necessary to make an enquiry among the users, sort the users in three categories indicated, and check the desired position in the graph with the actual position. There are means available by which we may attempt to bring these two points as close together as possible. This, of course, takes a lot of time, but after some time most of the profiles are stabilized.

Most users appreciate information retrieval systems which do not bother them with too much irrelevant information. They do not know how much they are losing in low recall. Though our users are satisfied with the service, we do feel that some improvement could be achieved in the way outlined.

We intend to sort the users into the groups indicating their orientation to either

Relevance (Rel) or
Recall (Rec) or
Compromise(R/R)

The recall figures would be calculated only in extreme cases e.g., where high recall is wanted but high relevance was achieved.

7.6 "Miss" and "Trash"

To evaluate the performance level of any information system, we may also use negative indicators, like "miss" (relevant not retrieved) or "trash" (irrelevant retrieved).

retrieved	relevant retrieved	irrelevant retrieved
not retrieved	relevant not retrieved	irrelevant not retrieved
	relevant	irrelevant

Fig. 18 Relevant/Irrelevant-Retrieved/Not Retrieved

One of these methods was used by R. A. Sprague, Jr. ("A Comparison of Systems for Selectively Disseminating Information," Bureau of Business Research, Graduate School of Business, Report No. 38. Bloomington: Indiana University, 1965). The equation

$$C = kM + T$$

attempts to express the cost (C) of a search to the user. "M" means "miss" or number of relevant not retrieved documents. The value of "M" is multiplied by the constant "k"; "k" is lower for those users which are relevance oriented (1) and high for recall oriented users (5). "T" stands for "trash" denoting the number of irrelevant retrieved documents.

As we need recall figures, we used for C evaluation the eight profiles for which we have established the recall figures. For each of these profiles we have determined the values of k, M, T and calculated C. We have determined the "k" by asking the user respective as to his relevance, recall or compromise orientation. We assigned the values 1, 3 or 5 respectively to this orientation to express it numerically. (We add relevance and recall figures to the tabulated "C" values, for comparison).

Rel.o. = Relevance oriented k = 1
R/R = Compromise k = 3
Rec.o. = Recall oriented k = 5

	Name	Profile	k	M	T	C	Relevance	Recall
A	Coldham, D.B.	100024	5	20	0	100	100	20
B	Karim, G.A.	100018	3	30	0	90	100	25
C	Dilger, W.	100021	5	14	1	71	88	33
D	Ross, G.A.	100023	3	32	10	106	62	33
E	Groves, T.K.	100009	3	18	23	77	61	67
F	Eder, W.E.	100026	1	57	106	163	52	67
G	Harrison, D.	100010	3	12	13	49	48	50
H	De Vries, G.	100018	1	2	11	13	20	67

Fig. 19 "C" Evaluation (December, 1969)

This table presents some interesting contribution to our inquiry into the performance of the system and of individual profiles (Figure 19).

Although the values of "k" range from 1 to 5, M from 2 through 57, T from 0 through 106 and C from 13 through 163, there is no indication that C by itself would be any indication of the users' satisfaction. All the users specified by A-H are essentially satisfied users. It seems to us that it will continue to be like this as long as the relevance-recall plot will show a reasonable configuration.

It appears that C alone is no absolute measure of system performance or users' satisfaction, but could be applied with some success to compare either individual profiles or systems, under comparable conditions; e.g., comparison of the profiles F (relevance 52, recall 67) with H (relevance 20, recall 67) of two relevance oriented users, would seem to be in favour of F because of higher relevance at an equal recall. But looking at the table we can readily see that C value for H is only thirteen (better) whereas for F it is 163 (worst of all) because this profile missed 57 abstracts and the trash is 106 records.

On the other hand the C value alone does not give us any idea of the relevance-recall values, e.g., the profile F is evaluated as the worst of the subset being examined. But, in spite of the 57 missed items, it was able to find two of each three relevant items in the data base and 52 of each hundred abstracts supplied were relevant.

We recommend to use both types of performance characteristics together: Thus "trash" would supplement relevance figures and "miss" would accompany recall figures. This would also provide for a better means to compare profile or system performance. It also allows us to make conclusions how to adjust the profile respective, if we add the orientation of the user either to relevance or recall.

e.g., evaluation

"Profile A (Rec.o.) Rel 100, T0; Rec 20, M20"

implies that for this particular profile an adjustment should be made to enhance his recall even at the expense of relevance, supposing the user considers the M too high.

On the other hand

"Profile E (R/R) Rel 61, T23; Rec 67, M18"

"Profile G (R/R) Rel 48, T13; Rec 50, M12" indicate that not too much could be improved for these medium oriented users.

The user with the following profile might require to improve his relevance:

"Profile H (Rel.o.) Rel 20, T11; Rec 67, M2"

but he does not because of the relatively low T.

The main advantage of this way of characterizing profiles is, that it not only gives the situation of the profile (relevance + recall), what it is losing (M) and what he is being disturbed with (T), but also the orientation of the user is indicated showing the direction of corrective steps. Systems could be characterized in a similar way.

7.7 Comparison of AND, WITH, ADJ

In order to ascertain the selectivity of AND, WITH,ADJ, logical connectors in practice, we have selected five profiles and we have conducted three searches after each other with the aforementioned logical connectors. Each time we have changed three search expressions of each of these profiles using the identical logical connectors. We have ascertained the number of hits for all of the five profiles with all three types of connectors. (See Figure 20.)

In choosing the profiles and the search expressions (the concepts in the original TEXT-PAC documentation) for this experimental run we were aware of the fact that this selection could affect the outcome of the experiment very considerably. We could select such groups of words which are very unlikely to lie close together or which, on the other hand, can only occur in a certain identical sequence. We did not adopt any of these extremes and we have chosen such words which can mostly be compounded with any of these logical connectors.

The results are shown in the following table:

Number of Profile	No. of Hits obtained		
	AND	WITH	ADJ
100001	413	299	239
100002	44	41	37
100017	110	107	101
100020	255	227	210
100025	251	226	198
Total	1,073	900	785
Job Time	5.21	5.02	4.19

Fig. 20 AND, WITH, ADJ and Hits Received

No general conclusions may be drawn from this table. If these profiles were searched against a very large data base, the number of hits would give the probability for these words to occur in a more or less tight connection. In our case they only indicate an example of how we can manipulate the search from a higher relevance to a higher expected recall (ADJ → AND).

It should be pointed out, that this tool must be used very carefully. There is no point in curbing the output by switching from AND to ADJ where such a combination has only a little chance to be found and there is no sense to look for two words apart from each other if they occur only in one specific sequence. Other, more appropriate, tools must be utilized in such a case.

7.8 Match Criteria 1 - 3

In order to see the effect of using match criterion greater than 1 we changed the match criterion to 2 and 3 respectively, on the header cards. We used 70 profiles and December, 1969 tape as the data base. If there was only one search expression, or two search expressions, in the profile we could go only as far with our match criteria respective.

As comfortable as the increasing the match criterion may seem to the user, (it requires only changing one digit on the header card), it is also the least precise: we make the hit dependent on the occurrence of two or more search expressions which:

1. might be relevant individually (either of them) but not collectively and so we lose relevant information (lower recall will result),

2. might give a false coordination (e.g., we are interested in both CON 1 PERT
CON 2 CAR$ or VEHICLE$. . . .
standing alone but we will get only information of PERT method in connection with car$ and information about car$ only in connection with PERT.)

The following table (Figure 21) illustrates how increasing match criterion reduces the number of hits and causes the number of profiles with no hits to rise.

	Match Criterion		
	1	2	3
Total number of hits	6301	2019	1406
Number of profiles with no hits	8	27	41

Fig. 21 Effect of MC on the Number of Hits

Increasing the match criterion may have varying effect with different profiles. Whereas with one profile (No. 100007) we decreased the number of hits 81 times (to 1.2 per cent) by setting MC = 2, in another case it was only 4.4 times (to 23 per cent). In this latter case we obtained 1942, 440, and 92 hits with M = 1, 2 or 3 respectively (No. 000017).

The effect of changing the match criterion depends largely on the quality of data base, on the profile words (if general or specific), on the logic used in search expressions (if loose or tight) and on the number of search expressions.

It may be concluded that a proper set-up of search expressions is preferable to increasing the match criterion.

7.9 Searching Titles, Subject Headings and Abstracts

This subject is not only of theoretical but very practical significance. Searching abstracts (or the entire text) is more elaborate and expensive. The question to be answered is whether this higher cost is reflected in a higher yield of information retrieval from a data base when searching from the full text instead of from titles or subject headings.

TEXT-PAC enables us to search the full text of individual records in the data base. We may also limit the search to any one print control or to a group of print controls. We may also exclude one or more print controls from being searched. This is not recommended, because limiting the search makes the system not to utilize the full capabilities of the system.

We did not have to set up our own experiment because three profiles have supplied the information required. The three profiles have the same profile words and logical connectors. They differ in that one of them is matched against titles, the second against subject headings and the third against abstracts. This is brought about by the CONTROL facility.

The wording of these profiles is as follows:

```
CON 1  COMPUTER$
CON 2  INFORMATION ADJ RETRIEVAL
CON 3  INFORMATION ADJ STORAGE
```

The results of running this profile in the three modifications are given in the table below.

Profile Number	PC Searched		Month, 1970 Jan	Feb	Mar	Apr	TOTAL	INDEX
000022	00$	Title	40	57	74	62	176	100
000023	09$, 60$	Subject Heading	49	47	109	98	256	145
000024	50$	Abstract	127	157	216	133	476	270

Fig. 22 Title, Subject and Abstract Searching

We can see from the above table that matching with abstracts of a given data base has yielded 2.7 times more hits than matching the same profile with titles only. With other profiles this result will be even more in favour of abstracts as abstracts dealing with "computers" and "information retrieval" always tend to have these words in title. Even searching in subject headings has given 1.45 times more hits than titles.

The outcome shown would be more clean-cut if we used more involved profiles which have only little chance to be matched in titles, and if we sought the whole record, not only the abstract.

In addition to higher yield, the full text searching, of course, allows us to move in a wider range of relevance -- recall trade-offs due to more exhaustive data base. An additional advantage is the possibility for the user to judge the relevance from the abstract.

8. STEP TIME OF SOME OF THE PROGRAMS

In evaluating any information system especially from the point of view of incurring costs, it is very important to study thoroughly all individual programs in terms of time necessary for their running under the conditions given or anticipated.

All main programs involved in running CIS sector of COMPENDEX (Selective Dissemination of Information, Current Information Service) may be subdivided in three groups, viz.:

1. Profile	Profile Update TRC001
	Profile Diagnostic TRC002
	Profile Print TRC003
2. Edit	360 Condensed Text Edit TRC260
	Edit Convert TRC210
	Edit Print TRC203
3. Search/Print	CIS Memory Load TRC010
	CIS Search TRC011
	CIS Answer Inversion TRC012
	CIS Disk Load TRC013
	CIS Print TRC014

1. Profile

In order to be able to determine the time involved in running the above programs, without CIS Print, we took the February/1969 data base and made seven successive runs with 10, 20, 30, 40, 50, 60 and 70 profiles respectively. The step times ascertained are given in Figure 23 illustrating the role of a given number of profiles on the step times for a constant data base (February 1969; 1,527 abstracts).

It was established that the profile programs (see above) are not time-consuming if the interest profiles are properly set up. Otherwise it is necessary to submit the corrections again. The profile programs play a minor part in calculating the computer time (Figure 24).

STEPS No. of Profiles	Step 1 001 Profile Update	Step 2 002 Profile Diagnostic	Step 3 003 Profile Print	Step 6 010 Memory Load	Step 7 011 CIS Search	Step 8 012 Answer Inversion + No Hits + Names	Step 9 013 Disk Load	TOTAL JOB
10	0.01	0.02	0.01	0.04	1.11	0.04	0.26	1.53
20	0.01	0.03	0.02	0.07	2.44	0.06	0.39	3.05
30	0.01	0.02	0.02	0.08	3.42	0.07	0.33	4.00
40	0.02	0.03	0.04	0.17	4.83	0.07	0.42	5.60
50	0.03	0.05	0.05	0.26	6.48	0.07	0.44	7.41
60	0.04	0.04	0.08	0.46	9.16	0.09	0.44	10.34
70	0.04	0.05	0.09	0.45	8.98	0.10	0.45	10.19

Fig. 23 Step Times (min), February, 1969;
1,527 abstracts

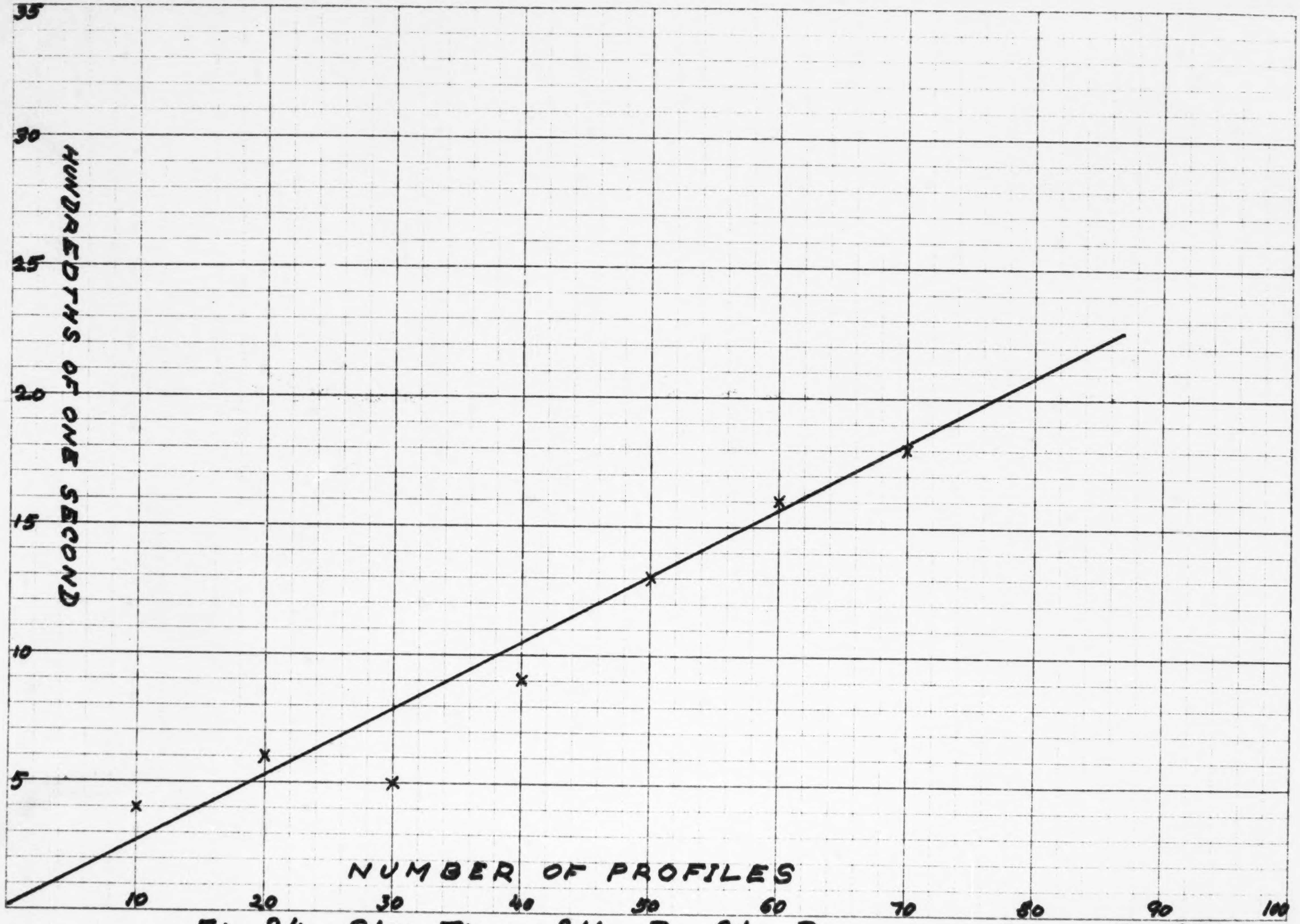


Fig. 24 - Step Time of the Profile Programs (10-70 Profiles)

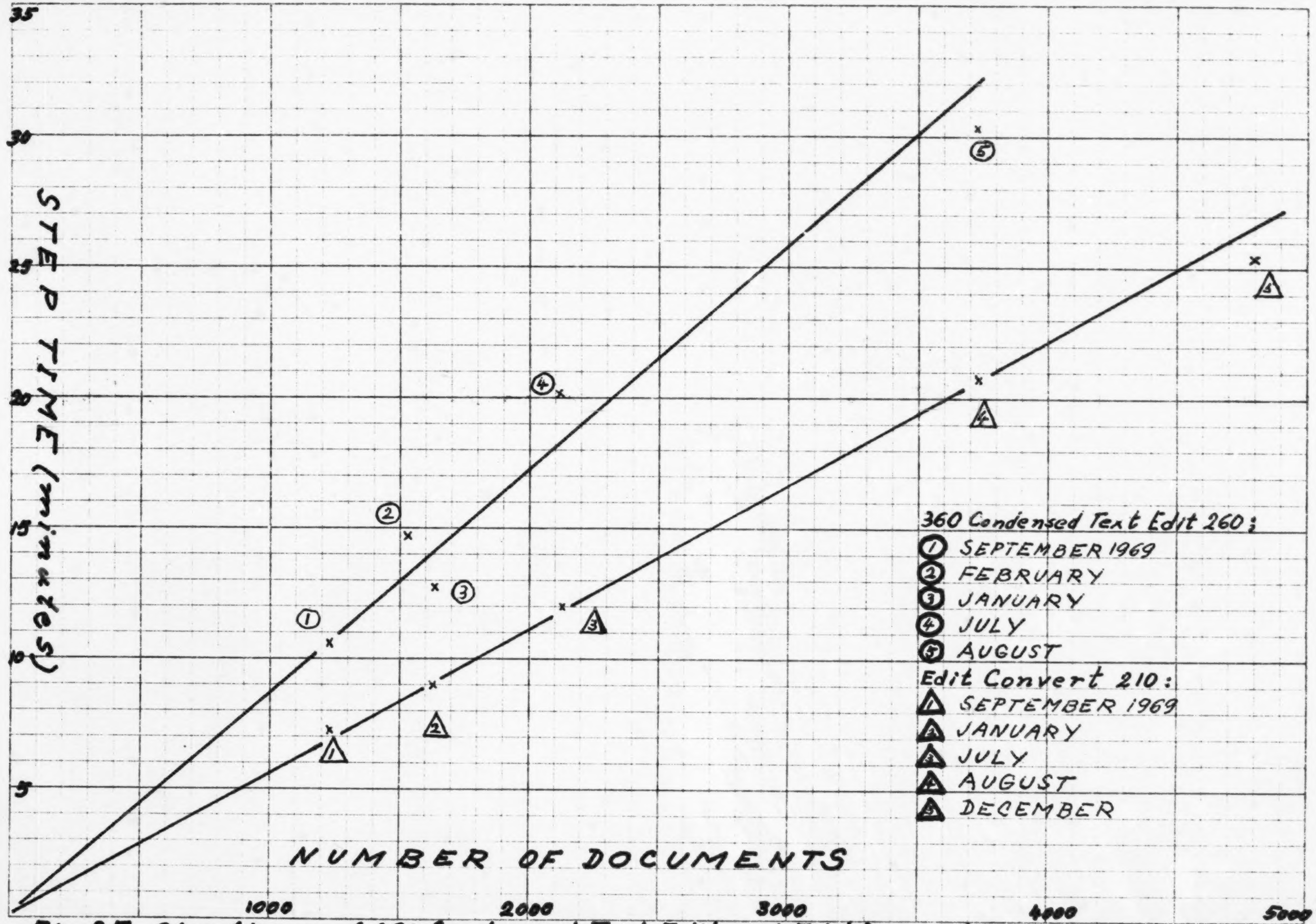


Fig. 25- Step times of 360 Condensed Text Edit and Edit Convert-Varying Number of Documents

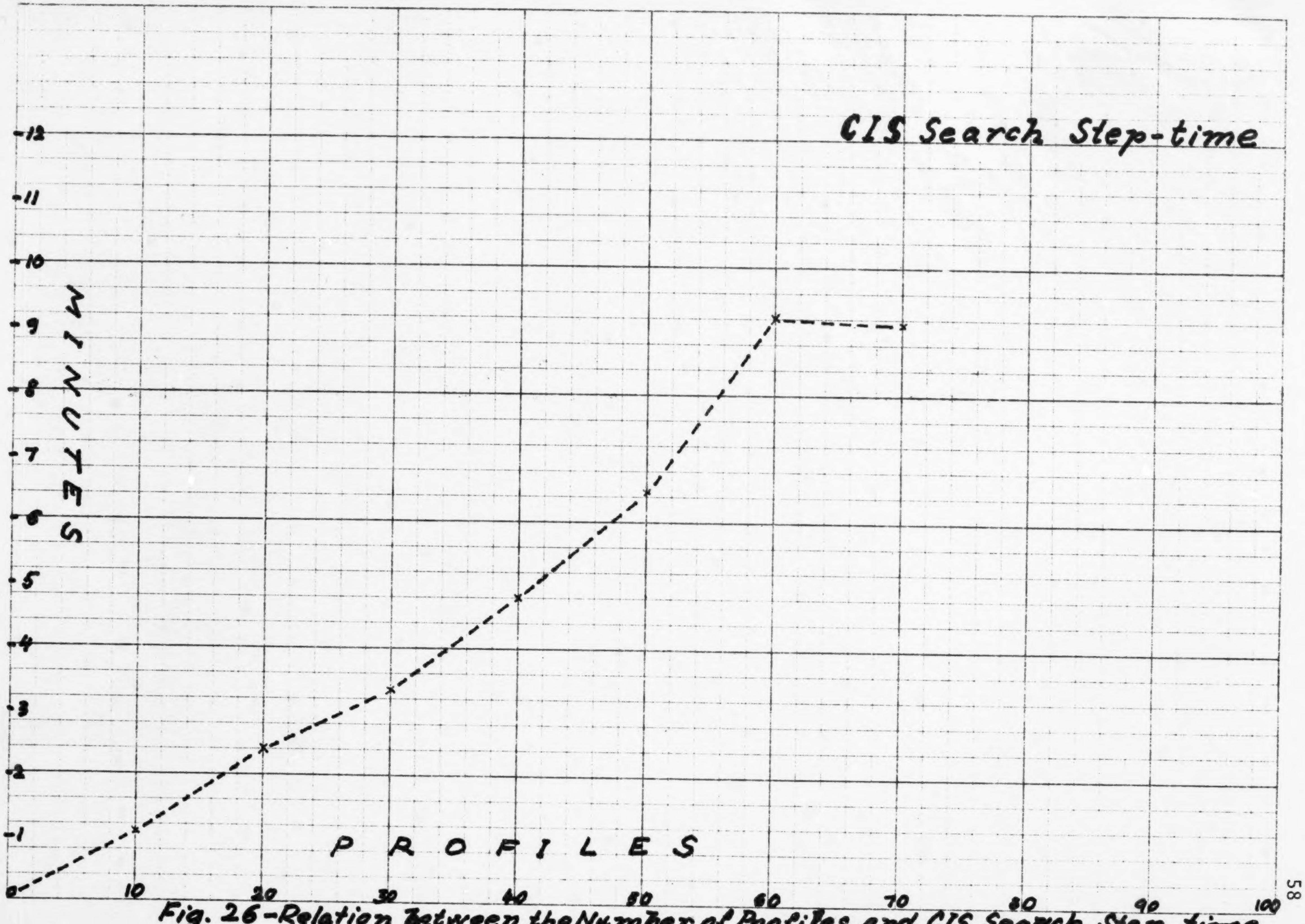


Fig. 26-Relation Between the Number of Profiles and CIS Search Step time

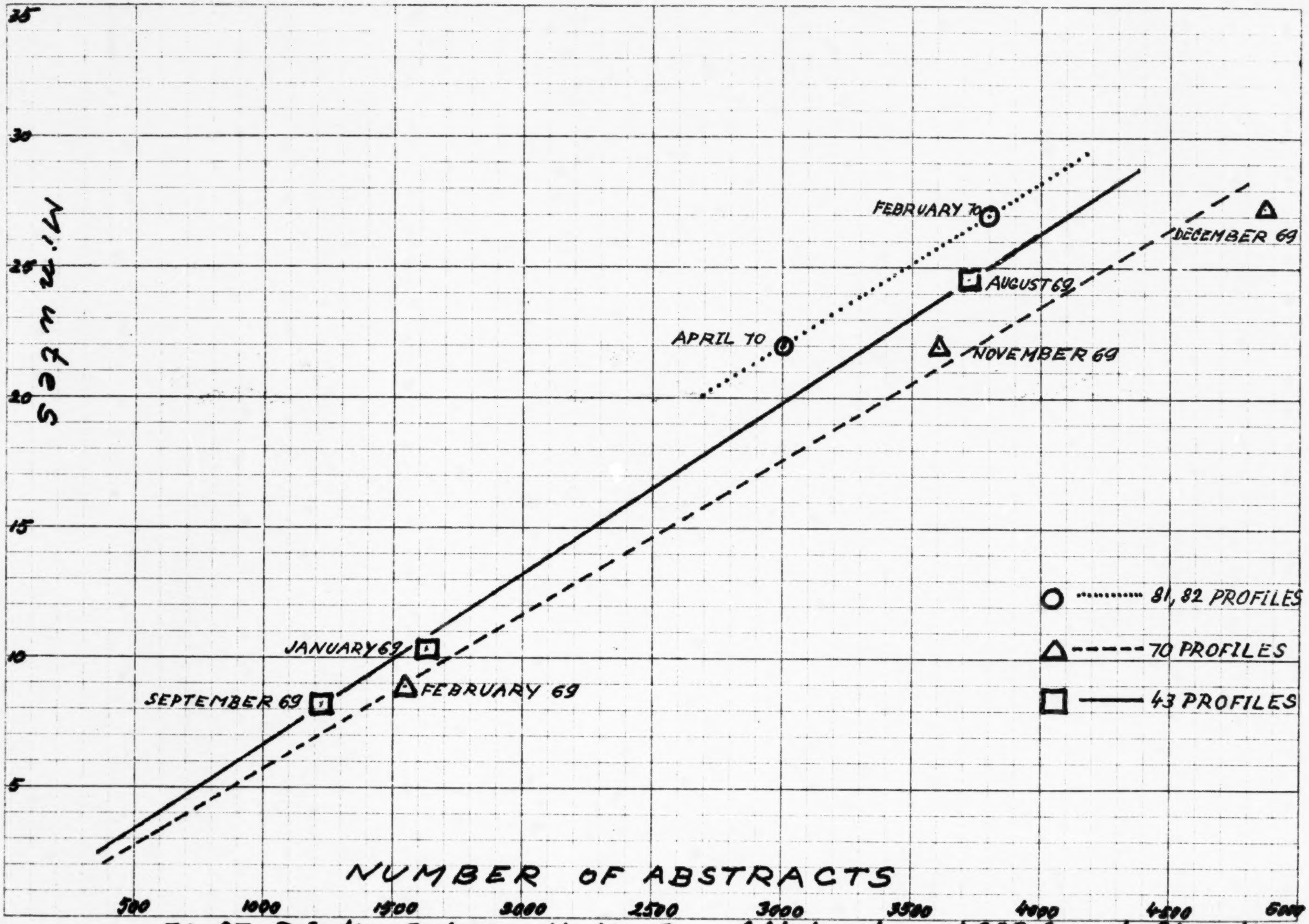


Fig. 27- Relation between the Number of Abstracts and CIS Search Step time

They take only a fraction of one second to run and the step time is explicitly related to the number of profiles and their structure.

2. Edit

All of the three Edit Programs are related to the size of the data base (number of documents) as far as their step times are concerned. To compare the monthly runs with each other and illustrate the impact of the number of documents on the step times of Edit Programs, we have compiled Figure 25. The graph was drawn clearly demonstrating the expected linear relationship of the time incurred and the number of documents. Two of these three programs 360 Condensed Text Edit 260 and Edit Convert 210, take a considerable share of time of the entire run (see Section 9).

3. Search/Print

Among the Search/Print programs the most time-consuming is the CIS Search TRC011. Logically, the step time should be affected by the number of profiles and by the number of records in the data base; the length of profiles and logic used are additional factors.

For a given data base the step time rises roughly proportionally when increasing the number of profiles (Figure 26). If the number of profiles increases over 100, two load modules will be needed to accommodate the profiles etc. As the data base will have to be run twice (successively against the first and second load module respectively), the step time necessary will grow up gradually (data base 4848 records, December, 1969):

Number of profiles	Step Time (mins)
70	28 (one module)
210	83 (three modules)

Fig. 28 Step times for 70 and 210 Profiles (CIS Search)

We have found out that the number of data base records has the same effect as the number of profiles (for a given number of load modules) (see Figure 27).

9. CALCULATION OF THE COST OF CURRENT INFORMATION SELECTION

There has been a dearth of published literature on the cost of information until recently. Though more information about this topic may be found now, the data published are not comparable among themselves. In evaluating any costs of information systems, we must remember that the cost of information must be always seen in the shadow of its value for the user(s). The relative cost of information is, therefore, hard to determine although the absolute costs may be well established; mainly because the value of one information may be zero for all other users except for one to whom it resolves a problem worth perhaps hundreds of thousands of dollars. But nobody can predict how many times the "right" information will find its "right" user in a system's environment.

Porter and Rudwick (Application of Cost-Effectiveness Analysis to EDP System Selection, MITRE Corp., Bedford, Mass. AD-667.522) distinguish, when selecting among alternative data processing systems, between "pivoting on constant effectiveness" and "pivoting on constant cost." In the first case one selects the system with the lowest total cost among systems with the same level of effectiveness; in the second case one adopts a system with highest level of effectiveness among systems not exceeding a specified total cost.

In discussing the economics of information systems, a great contribution was done by U. Hyslop (The Economics of Information Systems: Observations of Development Costs and Nature of the Market, American Society for Information Science Annual Meeting, Columbus, Ohio, 1968, Proceedings, Vol. 5, pp. 301-306). The author recognizes four major cost areas, namely (1) start-up costs, (2) operating costs, (3) continuing development and redesign costs, (4) marketing costs. Whereas the costs (1) should be subsidized, the costs (2) and (4) should be recovered from the users. A special attention is to be paid to the costs related to the continuing development and redesign, which should be also recovered from customers but some subsidy may appear necessary at the beginning.

The literature dealing with the costs of SDI systems is concentrated to the costs of operating the systems, but the figures are valid in a specific environment of different accounting methods, include only some of all incurring cost factors, are related to different data bases and numbers of users and so on. The Figure 41 reflects the fees charged for SDI services by various organizations giving some idea of the price but do not enable us to make any conclusion of their real costs and of the benefit to the user.

The opinions appraising the SDI systems cover the whole gamut extending from: ". . . least expensive, most efficient and most easily evaluated system to use as a base of information services" (Savage, T. R.: The Interpretation of SDI Data, American Documentation, 18, 4, October 1967, 242-246) to the opinion that SDI is relatively expensive in comparison with simple awareness methods such as circulation of secondary journals (Wente & Young, Operating Experience with NASA/SCAN, a Large Scale Selective Announcement Service, American Society for Information Science, Annual Meeting, Columbus, Ohio, 1968, Proceedings No. 5, pp. 217-223).

CIS ON CALGARY'S CAMPUS

In 1969 the Current Information Selection (CIS) was offered to the users on a free of charge basis. The system was run on an experimental scale, the objective having been the implementation of the COMPENDEX system, gathering the experience in the user-system interaction area and also making a calculation possible. The purpose of this calculation is double: (1) to elucidate what is the cost of operating this system and (2) what the charge of the users should be like. It is self-evident that any service which is of any value to anybody should be charged for, because otherwise there is no evidence of its usefulness. There are essentially three possible ways to raise sufficient funds for a service like that:

1. Totally from public resources (federal, provincial, municipal).
2. To bill the user for all the expenditures incurred.

3. To start with financial support and, once the system is operational, to charge the users partially or fully.

The third possibility seems to be the most justified. In this sense we have prepared a calculation which would provide for covering the costs of regular running the system once the pilot project is accomplished. Needless to say, there is no profit included in any of these considerations.

The variable factors mostly affecting any estimation of the costs are:

1. The number of abstracts, i. e. the size of the data base (the edit and search). We started with data base comprising over 1,000 documents, but the number has increased in December to over 4,800. We were assured by the Engineering Index that this is an average number on which to base and that further increase may be expected later on once the reformatting troubles in E.I. are overcome. Hence we took the December tape as representing an average data base at the present time.

2. The number of profiles. This is a hard predictable factor, since some of the users who participated on the pilot project may drop out, but there is a potential market for this service, especially if this service will be operating on a nationwide basis.

The higher the number of profiles, the more costly the CIS system, due partly to the step time of profile programs, but much more so due to the execution of search programs (the number of load modules) and printing more hits. An additional search editor represents further increase of costs. This increase in costs will be more than compensated by more revenue if the system of charging the users will be based on the number of profiles (and their length). Because the month of December, 1969 was run against 70 profiles we took this number for our calculation, and made a comparison with a 210 profile run taking further expansion into account.

3. The computer rates. There are no major changes to be expected in this area, either upwards nor downwards. We have to take the rate schedules effective this fiscal year.

4. Personnel costs. Two persons are foreseen to keep the system

running in the present extent on this campus.

5. There is a proportionate increase in the cost of material with the number of hits. This is represented mainly by the cost of the double response cards.

6. Overhead costs are included in the weights when calculating the hours of machine units.

7. Some additional system overhead amounting to 10 per cent of the salaries will reflect cost of correspondence, advertising, billing, accounting and mailing the information being disseminated.

The total monthly cost of the Current Information Selection is itemized in the following manner:

A. Computer Costs
B. 20 per cent of Computer cost reserve for the Dictionary and Statistics
C. Keypunching - Verifying

D. Consulting
E. Printing
F. Cost of the System (TEXT-PAC)
G. Material
    (g) Data Base (tapes)
    (gg) Tape Reel
    (ggg) Double Cards

H. Cost of Implementation

I. Salaries

J. Handling, Mailing, etc.

K. Other Overhead

The Costs of the CIS Mode (Selective Dissemination of Information) of COMPENDEX Service/Month (Month of December, 1969. Data base 4848 documents. 70 profiles.)

A. Computer Costs

Step time equals the CPU time

26 msec = 0.00043 min.

JOB TIME = CPU + (26 msec X I/O Waits)

Weights = Weight 1 = 1.575

Weight 2 = 0.154

= Weight 3 = 0.415

UNITS = (Weight 1 X CPU) + (Weight 2 X No. of Data Sets X JOB TIME) + (Weight 3 X $\frac{\text{region size}}{100}$ X JOB TIME)

COST = $\frac{\text{UNITS}}{100}$ X Rate/Hour = UNITS X 1.50

No.	Programs Involved in CIS
1	Profile Update TRC 001
2	Profile Diagnostic TRC 002
3	CIS Profile Print TRC 003
4	360 Condensed Text Edit TRC 262
5	Edit Convert TRC 210
6	Edit Print TRC 203
7	CIS Memory Load TRC 010
8	CIS Search TRC 011
9	CIS Answer Inversion TRC 012
10	NOHIT
11	NAMES
12	CIS Disk Load TRC 013
13	CIS Print TRC 014

Fig. 29 CIS Programs

No.	Data Sets	I/O Waits	Region (K)	Step Time (CPU)	Job Time	UNITS	$
1	5	309	52	0.04	0.17	0.23	0.35
2	5	285	52	0.11	0.23	0.40	0.60
3	5	1273	52	0.08	0.63	0.75	1.13
4	7	42	100	41.21	41.23	126.47	189.71
5	5	55	128	25.53	25.55	73.45	110.18
6	4	4679	52	2.33	4.34	7.28	10.92
7	11	647	128	0.44	0.72	2.30	3.45
8	5	472	128	28.16	28.36	81.25	121.86
9	7	267	128	0.19	0.30	0.78	1.17
10	5	55	52	0.01	0.03	0.05	0.08
11	5	188	52	0.02	0.10	0.13	0.20
12	7	11205	100	1.35	6.17	11.34	17.01
13	6	3103	100	1.61	2.94	6.48	9.72
TOTAL				101.08	110.77	310.91	466.38

Fig. 30 Computer Cost (70 profiles)

Carry Forward 466.38

Forward			$466.38
B. 20 per cent of Computer Costs			
Reserve for the Dictionary and Statistics			93.28
C. Keypunching - Verifying			
1 Hour/Month (on average)			7.67
D. Consulting			
1 Hour/Month			11.00
E. Printing			
Monthly rental of the printer $1040			
Discounted monthly rental $786			
Hours/Month (1 Shift) 176			
1 Hour $4.47			
3 Hours/Month			13.41
F. Cost of the System (TEXT-PAC)			000.00
G. Material			
(g) Data Base (tapes)		$500.00	
(gg) Tape Reel		25.00	
(ggg) Double Cards			
Price of 100,000 cards	$1,233.24		
Customs Duty and Sales Tax	422.78		
	$1,656.02		
Cost of 100 cards $1.66			
Cost of 6300 cards		104.58	
Total Material		$629.58	629.58
H. Cost of Implementation			
Cost of implementation is not included in the cost of the service			000.00
I. Salaries			
2 persons are considered at this stage			1,300.00
Carry Forward			2,521.32

Forward	$2,521.32
J. Handling, Mailing, etc.	
10 per cent of the salaries	130.00
K. Other Overhead	
All other overhead costs are included in A.	000.00
TOTAL MONTHLY COST OF CIS	$2,651.32

According to this calculation the cost of CIS service, provided 70 profiles are processed, would be:

Number of Profiles / Period	70 Profiles	One Profile
$ Monthly	2,651.32	37.88
$ Yearly	31,815.84	454.51

Fig. 31 Total Costs (70 Profiles)

Obviously, this price would be prohibitive for any private user. The solution to this problem lies in increasing the number of profiles to the amount which can be handled, after the implementation of the system, without increasing the personnel costs. This number of profiles depends on factors which were analysed in the Chapter Interaction System-Users.

For 70 profiles the cost would be

	$/Month	$/Year
Total cost	2,651.32	31,815.84
Per user	56.41	676.93
Per profile	37.88	454.51
Per search expression	5.35	64.14
Per word	1.07	12.88
Per hit	0.42	5.04

Fig. 32 Cost per User, Profile, Search Expression, Word and Hit (70 Profiles)

For this reason we have decided to perform a trial run with a considerably higher number of profiles. We did not have a sufficient number of profiles for this purpose and establishing of simulated profiles would have taken too much time. That is why we adopted the method as follows: we have taken the set of 70 profiles, placed them three times on the tape and obtained 210 profiles in this way. A minor change in program needed for proper numbering of profiles from 1 through 210 was all we had to do. We only were interested in cost evaluating and did not mind threefold repeating of the identical profiles. (As a check we have got exactly three times as much hits (18903) and no hits (24) as with the 70 profiles set.) In this manner we have been able not only to establish valid figures for 210 profiles, but we can estimate even further expansion by extrapolation. The results are given below:

No.	Programs Involved in CIS
1	Profile Update TRC 001
2	Profile Diagnostic TRC 002
3	CIS Profile Print TRC 003
4	360 Condensed Text Edit TRC 262
5	Edit Convert TRC 210
6	Edit Print TRC 203
7	CIS Memory Load TRC 010
8	CIS Search TRC 011
9	CIS Answer Inversion TRC 012
10	NOHIT
11	NAMES
12	CIS Disk Load TRC 013
13	CIS Print TRC 014

Fig. 33 CIS Programs in Fig. 34

No.	Data Sets	I/O Waits	Region (K)	Step Time (CPU)	Job Time	Units	$	Note (Related to 70 Profiles)
1	-	-	-	-	-	-	1.05	ESTIMATE (3X Higher)
2	-	-	-	-	[illegible]	-	1.80	ESTIMATE (3X Higher)
3	-	-	-	-	-	-	3.39	ESTIMATE (3X Higher)
4	7	42	100	41.21	41.23	126.47	189.71	Identical Data Base as for 70 Profiles
5	5	55	128	25.53	25.55	73.45	110.18	Identical Data Base as for 70 Profiles
6	4	4679	52	2.33	4.34	7.28	10.92	Identical Data Base as for 70 Profiles
7	9	1750	128	2.13	2.88	8.87	13.31	ACTUAL RUN
8	3	2909	128	83.00	84.25	214.40	321.60	ACTUAL RUN
9	4	664	128	0.80	1.09	2.51	3.77	ACTUAL RUN

Continued

(continued)

No.	Data Sets	I/O Waits	Region (K)	Step Time (CPU)	Job Time	Units	$	Note (Related to 70 Profiles)
10	2	83	128	0.25	0.29	0.63	0.95	ACTUAL RUN
11	2	459	128	0.26	0.46	3.30	4.95	ACTUAL RUN
12	-	-	-	-	-	-	20.00	Estimate for Identical Data Base (slightly rising with number of hits).
13	-	-	-	-	-	-	29.16	ESTIMATE 3 times more hits printed; cost 3x higher.
Total Computer Cost for 210 profiles							710.79	

Fig. 34 Computer Cost (210 Profiles)

A. Computer Costs		$710.79
B. 20 per cent of Computer Costs		142.16
C. Keypunching - Verifying		23.01
D. Consulting 1 hour per month		11.00
E. Printing Three times as much as with 70 profiles (see there) if we expect a proportionate increase of hits.		40.23
F. Cost of the System (TEXT-PAC)		000.00
G. Material		
(g) Data Base (tapes)	$500.00	
(gg) Tape Reel	25.00	
(ggg) Double Cards		
Cost of 100 cards $1.66		
Cost of 18,900 cards	313.74	
Total Material	$838.74	838.74
H. Cost of Implementation Cost of implementation is not included in the cost of service.		000.00
I. Salaries 2 persons		1,300.00
J. Handling, Mailing, etc. 10 per cent of the salaries		130.00
K. Other Overhead All other overhead costs are covered in A.		000.00
Total cost of a monthly run (210 profiles)		3,195.93

With 210 profiles

	$/Month	$/Year
Total Costs	3,195.93	38,351.16
Per user	22.83	273.94
Per profile	15.22	182.62
Per search expression	2.14	25.72
Per word	0.42	5.14
Per hit	0.17	2.03

Fig. 35 Cost per User, Profile, Search Expression, Word and Hit (210 Profiles)

In the above calculation we assume the same ratio profiles/users = 1.5/1 as has been with the 70 profiles runs, on average 7.1 search expressions/profile, 5 words/search expression, 35 words/profile, 53 words/user.

From the above figure it may be seen that increasing the number of profiles three times (from 70 to 210) or by 200 per cent, brings about only 20.54 per cent increase in the total cost whereas this cost is divided among 210 profiles. It substantiates our assumption that this is the way to make the cost per profile acceptable. The limits may be at about 300 profiles which can be handled by one search editor after the profiles had been verified in actual processing.

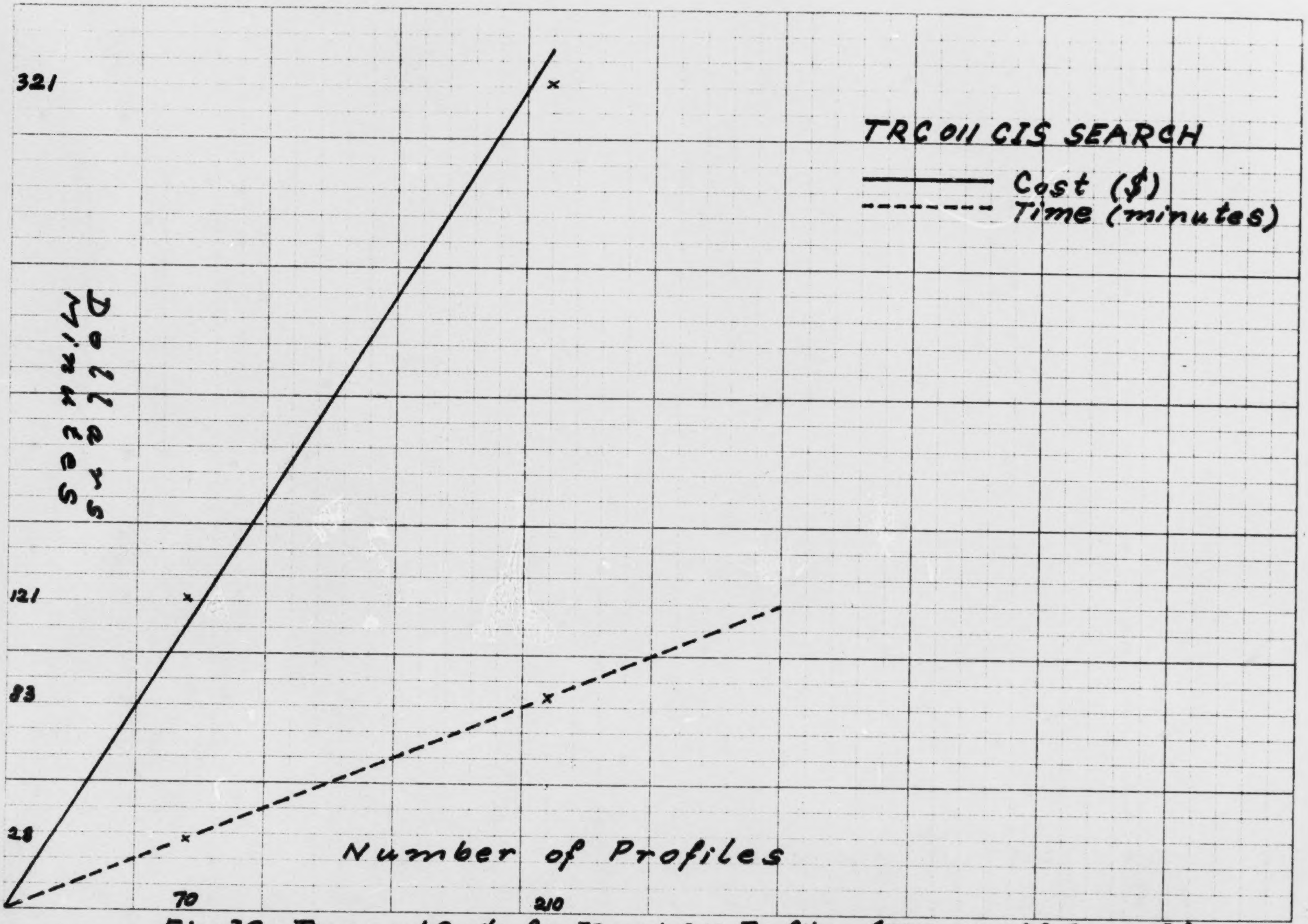


Fig. 36 - Time and Costs for 70 and 210 Profiles (TRC 011 CIS search)

COMPUTER COST

The Share of Individual Programs

	70 Profiles		210 Profiles	
No.	$	%	$	%
1	0.35	0.08	1.05	0.14
2	0.60	0.13	1.80	0.25
3	1.13	0.24	3.39	0.48
4	189.71	40.68	189.71	26.69
5	110.18	23.62	110.18	15.50
6	10.92	2.34	10.92	1.54
7	3.45	0.74	13.31	1.87
8	121.86	26.13	321.60	45.26
9	1.17	0.25	3.77	0.53
10	0.08	0.02	0.95	0.13
11	0.20	0.04	4.95	0.70
12	17.01	3.65	20.00	2.81
13	9.72	2.08	29.16	4.10
	466.38	100.00	710.79	100.00

Fig. 37 Cost of Individual Programs

In the above figure it is interesting to notice the declining share of the Condensed Text Edit (4) and Edit Convert (5) programs as converse to the rising cost of the Search program (8). Figure 36 reflects the rise of both the time (minutes of step time) and the cost ($) of the CIS Search TRC 011. Edit programs costs are fixed (4, 5, 6,). The share of individual costs in the total computer cost is illustrated in Figure 37 both for 70 and 210 profiles.

Figure 38 demonstrates rising costs of SDI/year and decreasing price/profile with an increasing number of profiles.

Figure 39 shows the percentage of costs A-K in the total cost both for 70 and 210 profiles.

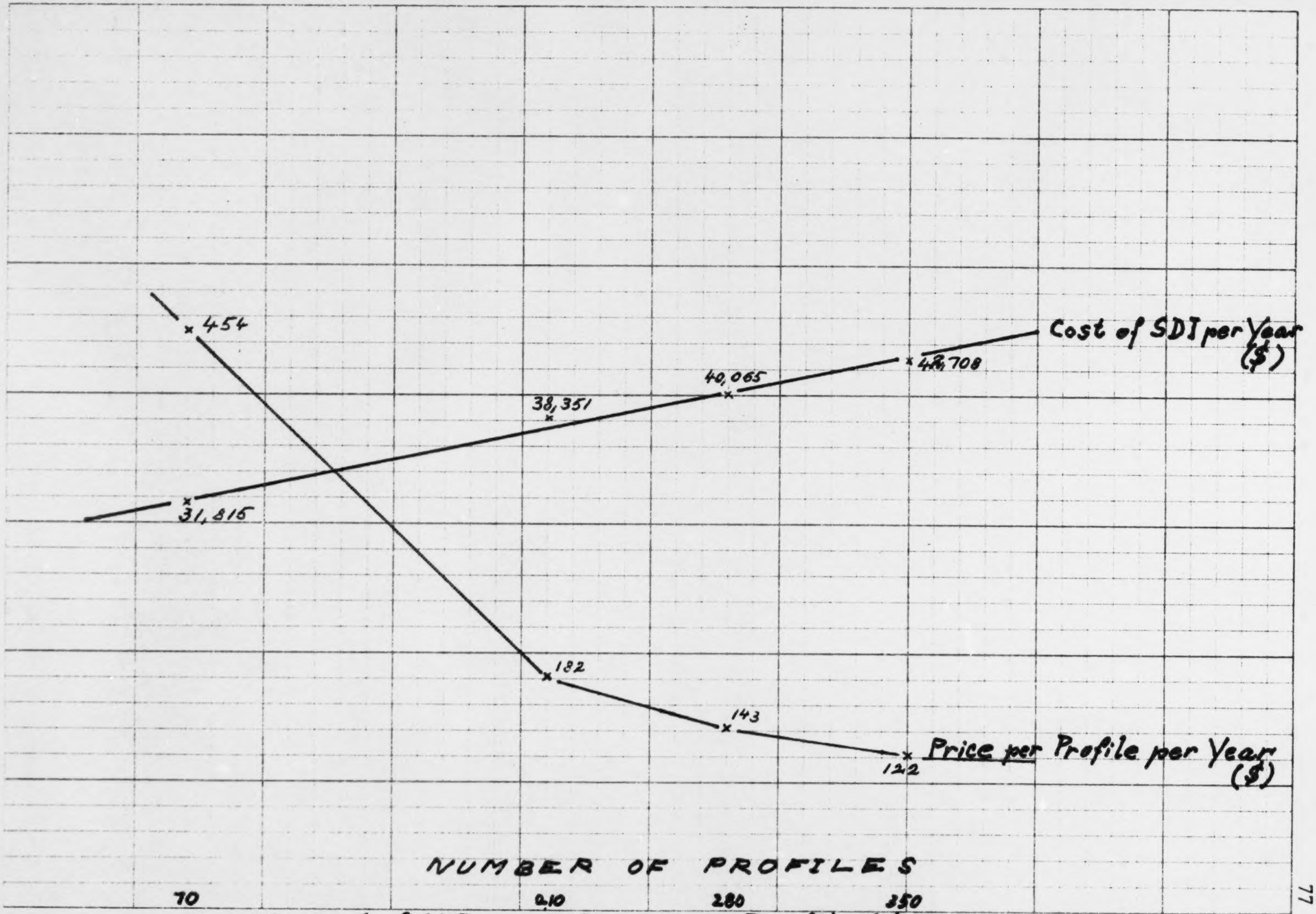


Fig. 38 - Cost of SDI and Price per Profile ($)

	Costs A through K (per month)	70 Profiles $	 %	210 Profiles $	 %
A.	Computer Cost	466.38	17.6	710.79	22.23
B.	20 per cent Reserve for Dictionary and Statistics	93.28	3.5	142.16	4.48
C.	Keypunching and Verifying	7.67	0.3	23.01	0.72
D.	Consulting	11.00	0.4	11.00	0.34
E.	Printing	13.41	0.5	40.23	1.26
F.	Cost of the System (TEXT-PAC)	000.00	0.0	000.00	0.00
G.	Material	629.58	23.8	838.74	26.23
H.	Cost of Implementation	000.00	0.0	000.00	0.00
I.	Salaries	1,300.00	49.0	1,300.00	40.67
J.	Handling, Mailing, etc.	130.00	4.9	130.00	4.07
K.	Other overhead	000.00	0.0	000.00	0.00
	TOTAL	2,651.32	100.0	3,195.93	100.00

Fig. 39 Cost A through K (70 and 210 Profiles)

Whereas some costs are fixed (salary), others are partially fixed and partially proportional (material, computer costs) others proportional (keypunching, printing).

If we anticipate, for the sake of simplification, a steady proportionate increase with the number of profiles (and we may do so because there is no progressively growing component), we obtain the following table:

	70	210	Estimate 280	Rough Estimate 350
Cost/Year	31,815.84	38,351.16	40,065	42,708
Price/Profile/Year	454.51	182.62	143	122

Fig. 40 Cost vs. Price per Profile
(70, 210, 280, 350 Profiles)

It may be concluded that, with increasing number of profiles and hence increasing number of hits (for an identical data base responsive to the profiles) we may expect slow increase in computer costs. This is largely due to the Search program. The total cost also slightly increases, mainly due to computer costs and material. The subscription price for profile decreases if the number of profiles is being held in a range which can be handled without increasing salaries. Number of profiles to be handled might be, after the start-up period, depending on their degree of sophistication and provided the search editor is relieved from clerical tasks, something up to 300. Under these circumstances, the price per profile could well be expected to drop below $140 (see Figures 38, 40).

The cost of one item on the magnetic tape delivered is as follows:

Number of abstracts	January	1,642
	February	1,527
	July	2,124
	August	3,738
	September	1,230
	October	3,673
	December	4,848
	TOTAL	18,782
	Average/Month	2,683
	Price/Month $	525
	Price/Item $	0.19

This price per item $0.19 will drop to $0.105 after we are supplied with 5,000 abstracts per month as promised

10. PRICING POLICY

Looking upon the table of what a user participating in diverse services is being charged, we may conclude that the amount is anything up to $225/user/year (Figure 41). Charging the user or his profile seems to be the most widespread method of billing. (This terminology assumes that one user may have several profiles each consisting of one or more search expressions, whereas sometimes user and profile are claimed to be identical.)

In COMPENDEX CIS mode it is appropriate to charge the user for his profiles (or search expressions), because the profile is a unit searched and so the number of profiles (or search expressions) is proportionate to the searching time. So is the number of words searched in any profile and a limit should be set on the number of words in a profile for a given rate to be charged. The rate is increased if the number of words is exceeded. But the user should be advised not to save words in defining his interest. Some discount should be allowed to users who submit their profiles (1) coded on sheets, (2) keypunched on cards. The user should save rather by submitting his profiles in form (1) or (2) than by leaving out words characterizing his special area of interest.

Some Information Centres charge the user according to the number of hits. This, in our mind, is a less appropriate criterion because.

1. if the charging for hits represents the charging for benefits from the systems, it need not be necessarily so; sometimes less hits contain more wanted information, cause less inconvenience in going through it; sometimes even no information is valuable information;

2. if charging for costs is involved, more hits mean more step time in the execution of CIS Answer Inversion, Disk Load, CIS Print 014, and more printing; but these steps are not time-consuming and do not influence the cost too much. Furthermore, should the user wish to save in limiting the number of printed hits, he may do so with systems using weights and ordering the hits accordingly, but he may miss the useful information right behind the limit set by himself.

A fair approach would be to charge for relevant information, but this is not feasible.

The pricing policy for COMPENDEX service should be, in our mind, based on the following principles:

1. The costs are partly subsidized but increasingly covered by revenue.

2. No profit is involved.

3. The rates should not cause the charge for the service to be restrictive (prohibitive).

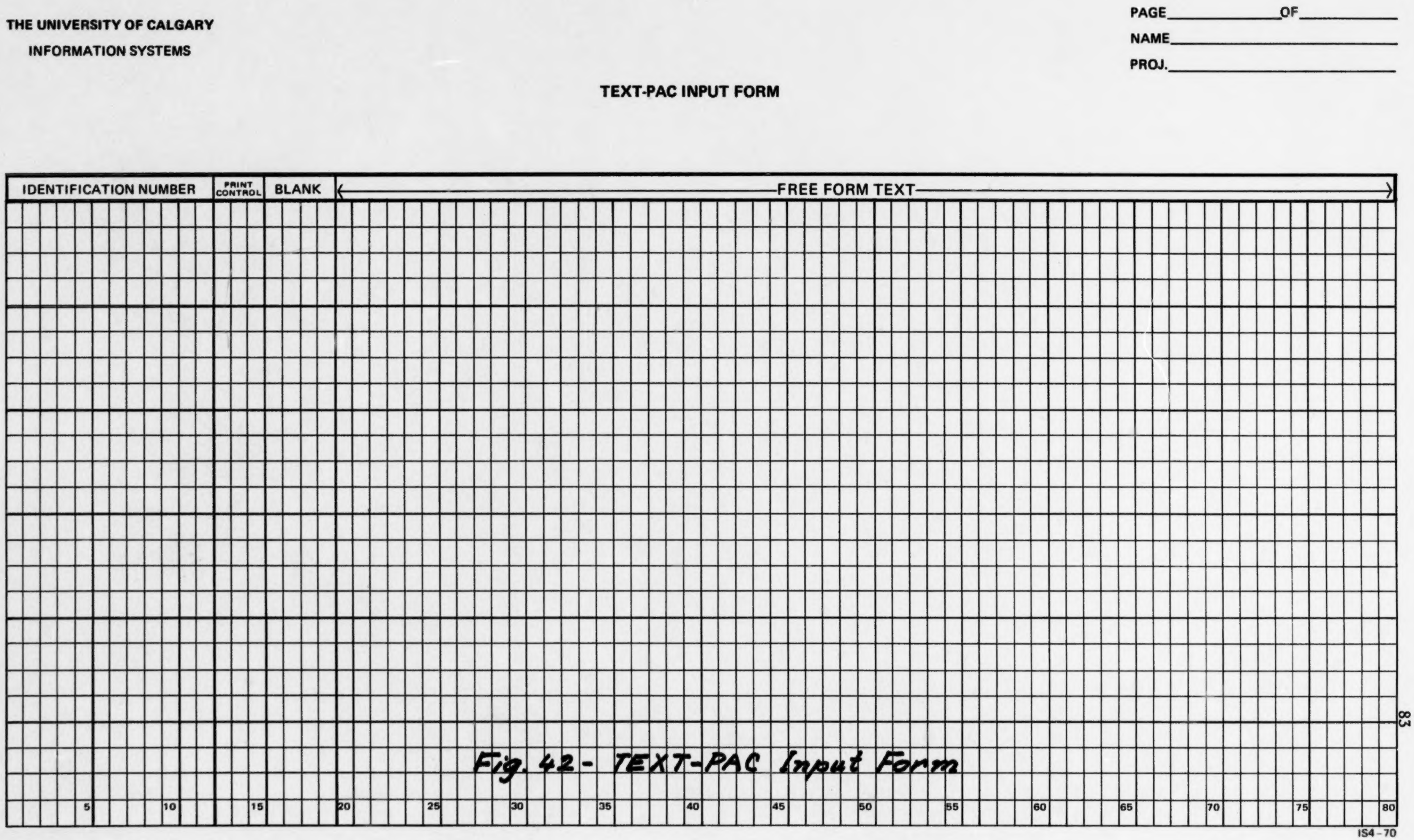
4. The rates should have an impact on the user in accordance with his usage of the system (increasing the costs of the system) rather than with his benefit from the service which is hard to assess.

5. The pricing system should be simple so as not to involve much clerical work and overhead costs.

6. The pricing system should be easily intelligible to the user. This matter is of prime concern to the user and he is not willing to

SDI System	Charge	Note
PANDEX (CCM Information Corporation N.Y.)	(1) Per Profile $150/profile<60 terms/year + handling + mailing + $3/term if>60 terms/profile + $0.03/citation if >30 citations OR (2) Running User's Own SDI Program $10,000.00/year + $50/hour computer time + keypunching + handling + mailing	Letter of January 16, 1970.
CHEMICAL TITLES AND ISI TAPES (National Science Library, Ottawa)	$100/profile<60 terms/year + $100 if>60<160 terms/profile This nominal charge does not cover the total cost.	NSL Newsletter October 18, 1968.
U.S. AIR FORCE	$15/user/year	
DAY U.S., NASA	$100-$150/user/year	Selective Dissemination of Information AD-674168
UNIVERSITY OF GEORGIA	$120/year	
NASA/SCAN	$18.50/user/year	
U.S. ARMY ECOM	$58/user/year	
DOW CHEMICAL	$65/user/year	
INDIANA UNIVERSITY	$145-$206/user/year	Experimental
AMES LAB. USAEC	$150-/user/year	
SUNY TIDB	$225/user/year	
SCIENTIFIC DOCUM. CENTRE	$0.05 per hit	
NATIONAL CANCER INSTIT.	$0.088 per hit	
COMPENDEX (AIRA: The University of Calgary, Information Systems	$10.00 $100/profile<40 terms/year	Token fee until July 1, 1970. Tentatively after July 1, 1970

Fig. 41 SDI Price



THE UNIVERSITY OF CALGARY
INFORMATION SYSTEMS

PAGE________OF________
NAME________________
PROJ.________________

TEXT-PAC INPUT FORM

IDENTIFICATION NUMBER
PRINT CONTROL
BLANK
FREE FORM TEXT
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
IS4-70

Fig. 42- TEXT-PAC Input Form

study any comprehensive pricing instructions.

7. The billing should be annually in advance to facilitate the budgeting of the system.

Alberta Information Retrieval Association charges $100.00/year/ profile in the COMPENDEX service, provided the profile does not contain more than 40 terms. Any additional 10 terms would be $20.00.

11. INPUT TO TEXT-PAC OTHER THAN COMPENDEX

Within the framework of this development some attention was also paid to the use of TEXT-PAC for a data base other than COMPENDEX. Some interest arose on this campus to put in some information in free form text and have a capability of full text searching. An Original Text Input Form was, therefore, designed (Figure 42) with comments and a small trial batch of 20 cards successfully edited. The following is an explanation to the input form.

When preparing the full text source document (e.g. an abstract) we always indicate 12 characters of the identification number. The first three characters of this number must be alphabetic. Print control designates the different data elements within an identification number. The first character must always be numeric. We have adopted the print controls as follows:

00ƀ	Title
10ƀ	Identification number
201	First author
202-299	Second to 99th author
4ƀZ	Source
50ƀ	Abstract
60ƀ	Subject heading, subheading
650-699	Access words

Each line in the input form must begin with identification number and a print control, otherwise an error message will result.

Full text begins in the column 20. The following rules are to be observed:

1. The maximum number of words per line is 16.
2. Any of the print controls may contain as many as 54 lines.
3. Maximum word length is 20 characters for comparing.
4. Initial capitalization is indicated by one "at" sign (@).
5. All letters in upper case are indicated by double "at" sign (@@) at the beginning of any particular word.
6. Punctuation is coded as the last character of the word (without blank).
7. Spacing e.g. between heading, subheading, etc., is brought about by number sign #, which is attached as the last character of the word (without blank).
8. The end of a sentence is assumed, if
    (a) a period, question mark or an exclamation point is followed by two consecutive blanks,
    (b) any special character is followed by two consecutive blanks,
    (c) a period, question mark or an exclamation point is placed in the column 79 and is followed by a blank in column 80,
    (d) a period, question mark or an exclamation point appears in the column 80.
9. Three consecutive blanks on a line mean termination of the text on this line (on this punched card).

If there are any errors they must be eliminated by correction cards. The maximum number of words permitted per line is the same as in the input cards (16) and so is the number of lines per print control (54).

The correction code (columns 23-24) varies according to the nature of the correction desired:

DT	Delete entire data item headed by this identification number.
D*	Delete from this print control.

DC	Delete just this print control.
RL	Replace a line.
AL	Add a line following the line number specified.
DL	Delete a line.
RW	Replace a word. This card deletes and adds words at the same time. The replacement may exceed one line or several lines, though the words to be replaced must be contiguous on the line specified.
DW	In contrast to RW, by means of DW only the words within the specified line may be deleted.
AW	The words to be added are specified in the columns 29-80 and the additions begin right following the word number indicated in columns 18-19.

12. SOME LIMITING FACTORS IN THE TEXT-PAC SYSTEM

The match criterion:	1-9 in CIS, 1-19 in RETRO
The query word length:	Maximum 38 characters Internal truncation to 20 characters searchable
Selective masking:	Maximum 6$ - 6 characters
Unconditional masking:	Matches all words to a total of 20 characters
CONTROL, NOT-CONTROL:	Up to 7 print controls per question word permitted
AND:	Connects maximum 15 query words
Back referencing to logical symbols:	Maximum 15 times
Levels of back-referencing:	Maximum 3
User's last name:	Maximum 20 characters
Logical symbol:	5 characters (first character alphabetic)
Length of a logic level:	Maximum 10 cards (9 continuation cards) Maximum 15 logical symbols
TEXT-PAC input form:	Maximum number of words/line: 16 Maximum number of lines/print control: 54 Maximum number of characters/word: 20

13. CONCLUSIONS

COMPENDEX service has established itself on this campus and is gaining ground all across Canada. This is because of its renowned data base and the full text processing capability the superiority of which has been demonstrated. Users belong to all areas of engineering at universities, in industry and other organizations, in production, research, administration and education.

The communication with users is person to person, by phone or in writing on Calgary campus. Users outside of campus are served by AIRA. At the present time no advertising is being done on this campus. The number of AIRA customers is steadily increasing. In July, 1970, 106 profiles were run.

The performance of the system is quite reasonable. The relevance on Calgary campus for the months preceding December, 1969, was 44 per cent, in December, 1969, it was 60 per cent (AIRA 40 per cent). In January, February and March the output was manually scanned and the relevance has risen to 76 per cent, 73 per cent, and 69 per cent respectively. Not all feedback from the users is available as yet, but at present the relevance for April, May, June, July, 1969, is shown to be 47 per cent, 54 per cent, 55 per cent and 68 per cent respectively. While enhancing relevance, you may considerably lower recall. It depends on the knowledgeability of the scanning person in each particular profile. By a double check we have found that in one profile as many as 10 per cent of the screened out material might be considered relevant. This costly measure should be applied to relevance-oriented users only. Although no generally valid rule can be stated, it appears that relevance over 70 per cent can be reached with systems operating at 60 per cent and below.

Analysis of relevance has shown, that users do not label the output "relevant" or "irrelevant" properly. In their feedback, they tend, sometimes, to express their negative attitude to the information by labelling it as "irrelevant." It has come out during this work, that the most powerful means to improve the relevance is to find the right degree of specificity and exhaustivity in formulating profiles.

It should be pointed out, that manual scanning should in no way make up for a faulty profile set-up. It should only obviate errors due to typing, coding, punching, computer, program, ambiguous terms, and cases where the profile is all right nevertheless some irrelevance occurs anyway.

The ways to monitor relevance were shown to be in the logic used and in the proper degree of specificity, exhaustivity. First of all, however, one must determine the desired proportion between relevance and recall for any particular user.

A method for determining recall was described and practically verified. It has proven as a useful tool to complete the picture offered by relevance, both for a profile and the system as a whole. The recall was found to be in reasonable limits and it was demonstrated in a relevance/recall graph indicating roughly the region our system is operating in.

Though only eight profiles were assessed regarding their recall values, the results may be regarded fairly representative, because nearly 7,000 abstracts were virtually scanned and the profiles taken reflect all levels of relevance from 20 - 100 per cent. The inverse relationship between relevance and recall was substantiated.

The analysis of recall failures has underlined a need for proper formulation of the profile, very much like relevance. The search expressions in the profiles were either too exhaustive, or too specific terms were used, or not all possible approaches were attempted to formulate the need, or not all synonyms were specified, or the logic was too restrictive (most frequently). Here, the same applies as was stated for relevance: we must be aware which direction we want to move.

We have seen in evaluation of our system that relevance with recall is much better than relevance figures only to characterize a profile or a system. It was also demonstrated that the so-called "miss" (relevant information not retrieved) and "trash" (irrelevant information retrieved) are a valuable supplement of relevance and recall values. So is the orientation of a user.

It was illustrated how using logical connectors AND, WITH, ADJ can affect the number of hits. It can serve as one of very efficient means to monitor the relevance-recall relation.

On the other hand, it was shown that increasing the match criterion may be very harmful as far as recall is concerned.

The merits of full text processing were demonstrated by comparing title, subject heading and abstract searches.

The programs in CIS mode of TEXT-PAC are essentially profile programs, edit programs and search-print programs. The first named do not play any important part in terms of the step times. Step times of the edit programs are directly proportional to the number of records. Search program step time is directly proportional to the number of records and profiles and rises gradually with the number of memory loads (approximately 100 profiles).

The cost of running the system was calculated first for 70 profiles and 4848 records. This cost appeared to be prohibitive. We have analyzed the nature of individual cost items. The only remedy was to increase the number of profiles, as there was no item progressively increasing. Only the cost of the CIS Search Program, which rises proportionally with the number of records and profiles, steps up with the number of memory loads. The number of profiles must not exceed the amount which can be operated by the existing personnel. Under the circumstances the total cost/year should rise from 70 to 210 and 280 profiles from $31,800 to $38,300 and $40,000 respectively. The price per profile/year would decrease like this: $454, $182, $143.

The following recommendations seem to apply to the present status of implementation:

- Evaluation of the system is not a one-time job but a continuous one. Whereas it is impossible to ask the user to judge the recall, his views regarding e.g. completeness of coverage, quality of abstracts and their terminology, are invaluable.
- We have to continue checking the data base for misspellings and other errors. In full text processing this is especially important.
- Training search editors and users should be continued.

Importance of feedback should be pointed out at these courses. The instruction should include, first of all, correct completing of the COMPENDEX Profile Submission Form. This is the fundamental document in the communication user-system.

- The users should be classified from the beginning as to their orientation towards either a high relevance or recall or medium. This would facilitate monitoring their output.

- "Word Frequency" Listings (or "Dictionaries") are valuable means for correct profile formulating. They are a bridge between the abstracter's and search editor's vocabulary.

- After the first change (in the printing program) enabling us to order hard copies by means of the response card, the next advisable changes would be:

- change providing for an automatic relevance calculation
- change to indicate the search expression which has caused a hit
- automatic profile adjustment would be of great benefit, but is very sophisticated with the logic involved.